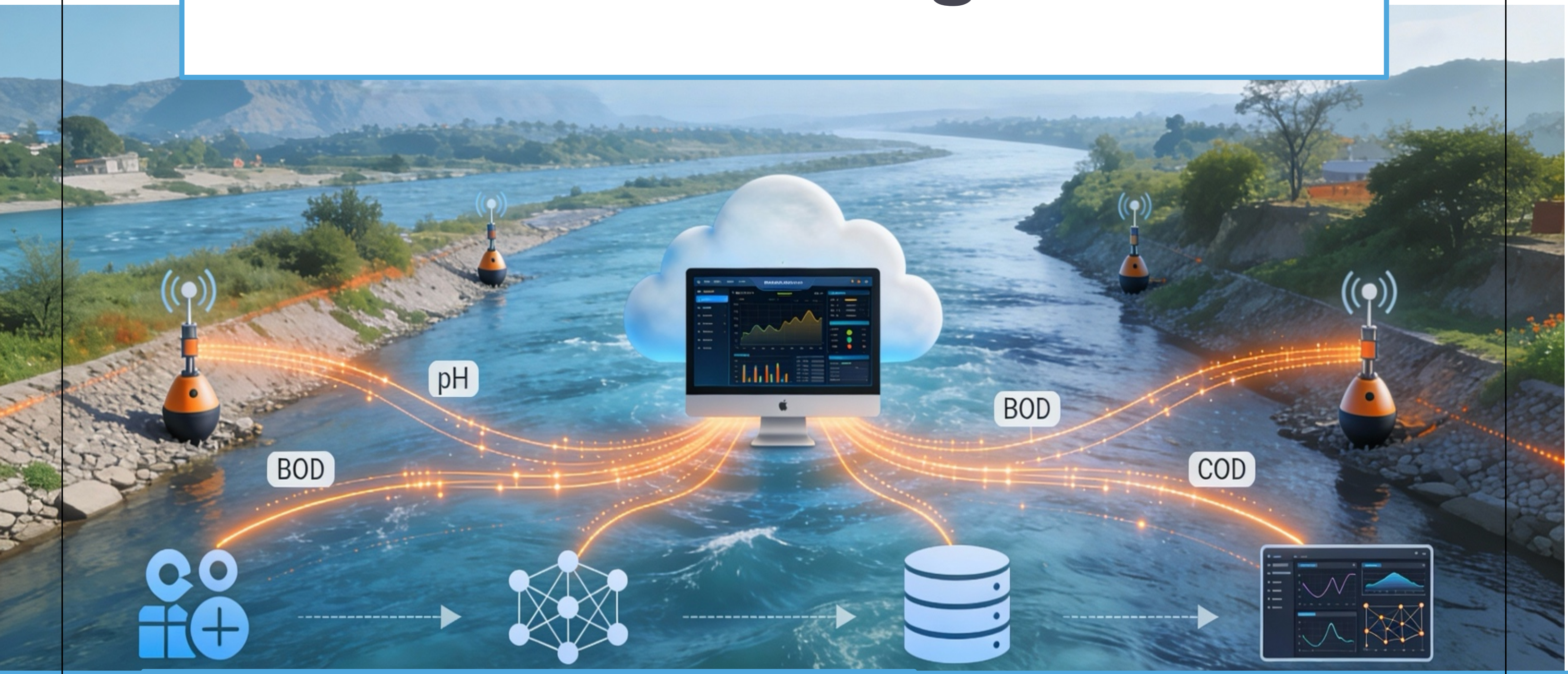
Ganga Water Quality Monitoring: Data Management Methodologies



March 2026

Expert Author: Dr. Sonja Behmel

International Cooperation with National Mission for Clean Ganga Project: Strengthening Quality Infrastructure for Water Monitoring of the River Ganga II

Physikalisch-Technische Bundesanstalt
National Metrology Institute

NAMAMI GANGE

german cooperation
DEUTSCHE ZUSAMMENARBEIT

Note:

This document was produced in the context of the project: Strengthening Quality Infrastructure for Water Monitoring of the Ganges River II for the National Metrology Institute PTB. All contents are related specifically to this project and should not be reproduced or taken out of this specific context. In addition, it must be noted that the content is built on previous work completed within the project, made available only to project partners. For further information, please contact PTB.

List of abbreviations

CPCB: Central Pollution Control Board
GIS: Geographical Information System
IFMLD: Integrative Field Data, Metadata, Laboratory Data and Environmental Data Management System
IWM: Integrated watershed management
LIMS: Laboratory Information Management System
NMCG: National Mission of Clean Ganga
PTB: National Institute of Metrology
SMCG: State Mission of Clean Ganga
WQ: Water quality
WQMP: Water quality monitoring program

Table of Contents

1. General Introduction

1.2 Methodologies for managing a Water Quality Monitoring Program

Declining water quality (WQ) has become a global issue of concern. Increasingly, integrated watershed management (IWM) is used to reform water governance in line with sustainable development, and to achieve targets for preventing and managing water pollution. IWM implies that all the stakeholders within a watershed (including the public) should be involved to make joint decisions and take actions to protect the resource. Two of the main challenges posed by IWM are (1) getting a reliable assessment of surface WQ in a given watershed through water quality monitoring programs (WQMPs), so that decision-makers can understand and use the information to support management activities, and (2) involving stakeholders in the IWM process and implementation of WQMPs so that they can take part in decision-making, understand the stakes, and integrate this knowledge into decisions and actions to protect the resource.

To start, we must distinguish the concepts of WQ assessment, study, surveillance and monitoring. In this context, “WQ assessment” is a generic term referring to any activity that aims to evaluate and estimate WQ. A “WQ study” implies an intensive deployment of means within a limited time frame to measure and observe the quality of the environment with one (or several) precise objective(s). A study therefore does not aim to monitor long-term changes in WQ over space and time within a watershed. “Surveillance” implies taking measurements and observations of WQ for management purposes, such as the surveillance of WQ at the raw water intake of a treatment station for drinking water. “Monitoring” implies a long-term, spatially distributed, standardized surveillance and quality assessment of all the WQ assessment activities and includes planning sampling strategies based on knowledge of WQ needs, field work, data handling, storage, and analysis, and the subsequent dissemination and use of the information that is generated (Behmel et al. 2016) (Figure 1). Only when these elements are included it is possible to speak of a full-fledged WQMP.

Figure 1: Steps of a WQMP. Adapted from Behmel (2010)

The objective of the present document is to provide the following for the Ganga River WQMP:

1. A methodology for quality control of data acquisition within WQMP, with a focus on managing an existing WQMP and field work.
2. A methodology for quality control of data and meta data management practices within WQMPs.
3. A methodology for quality control and processes for data validation and analysis within WQMPs.
4. A methodology for quality control and data sharing.

The methodologies are based a Data Management Needs Analysis of Surface Water Quality Monitoring and the recommendations issued from previous activities within the project. These can be found in Appendix B.

Nota bene: The present document does not include the methodologies to plan and optimize a Water Quality Monitoring Program; it focuses only on the management aspects specified above.

1.3 Definitions

For optimal understanding of the following sections, please refer to these definitions:

Database
A database is an organized collection of structured data, typically stored electronically and managed by a database management system (DBMS). This system, along with associated applications, is called a database system or database. Most modern databases model data in tables with rows and columns for efficient processing and querying using SQL.
(Source: What Is a Database | Oracle Canada).

Data
Information, especially facts or numbers, collected to be examined, considered, and used to help decision-making, or information in an electronic form that can be stored and used by a computer.
(Source: Cambridge Dictionary online, retrieved 2024-02-19).

Metadata
Information that is given to describe or help in using other information.
(Source: Cambridge Dictionary online, retrieved 2024-02-19)

Real-time Monitoring Data Cloud:
A type of system whose main purpose is to collect data from on-site hardware, networks, etc., and which allows the results to be visualized on user interfaces.
(Source: www.techtarget.com).

Laboratory Information Management System (LIMS)
A centralized platform to manage and organize data related to various laboratory processes, and whose main purpose is to streamline and optimize these processes.
(Source: Laboratory information management system – Wikipedia).

Geographic Information System (GIS):
A computer-based tool that examines spatial relationships, patterns and trends in geography, and whose main purpose is to store, analyze, and visualize data for geographic positions on Earth’s surface.

Project Management System:
A type of system whose main purpose is to provide a structured approach to managing operations within a company.

Integrative Field Data, Metadata, Laboratory Data and Environmental Data Management System (IFMLD):
An IFMLD is a centralized information system for the environmental data related data including the actual water monitoring data.
Its main purpose is to bring all relevant monitoring information into one central place and make it consistently usable from planning to reporting and publication.
It covers the full workflow of a Water Quality Monitoring Program (WQMP): the rationale and documentation of the program, fieldwork, sampling context (where, when, how, and under which conditions a sample was taken), laboratory results, and laboratory tracking information (often derived from a LIMS), together with the required metadata (e.g. sites, parameters, units, methods, instruments, quality flags, and versioning).
This system helps avoid transcription errors in environmental data, and includes communication features, procedures for streamlined data use, etc. It provides maximum information for data analysis and interpretation to achieve the most reliable results. It also supports controlled data sharing and publication, based on defined conditions, roles, and protocols.

2. Methodology for quality control of data acquisition within WQMP

Managing existing WQMP and field work

2.1 Introduction

The present chapter presumes that the sampling sites have already been selected and validated, the water quality parameters have been determined, the sampling strategy is defined, the logistics have been prepared, the sampling devices have been identified, and the laboratories have been determined.

The proposed methodology of this chapter (data acquisition) focuses on the following aspects:

i. Managing the information about the rationale of the WQMP, in order to plan and execute field work.
ii. Execution of the field work.
iii. Reporting on the field work.

Data acquisition within a WQMP can start well before the actual sampling starts. This type of data acquisition (or collection), includes, but is not limited to, collecting data about the watershed, land use, descriptions of waterbodies, hydrology, climatology, stakeholders, existing water quality and quantity data, topography, etc.

In order to implement optimal quality control, it is essential to use a database to properly document and manage information about the rationale of the WQMP, how the field work is executed, and any field work execution notes.

2.2 Current situation of the Ganga WQMP

For the Ganga WQMP, the current situation for managing information about the rationale of the WQMP is presented in Table 1, while the current situation for managing field data (sampling contexts) is presented in Table 2. It should be noted that there is currently no integrative field data, metadata and environmental data management system (IFMLD). Please refer to Appendix B for the full context on the content of these two tables.

Table 1: Managing WQMP rationale information of Ganga WQMP

Criteria	Status	Notes (based on Fieldwork observations)
WQM objectives	Documented, but not within a data management system	WQM objectives may not be fully aligned with sampling strategies.
Methodology for identifying monitoring objectives and translating them into sampling site selection, water quality parameters, and sampling frequency and recurrence.	Not readily available.	Need for a comprehensive review of the current WQMP, including the sampling site network, site validation, and sampling strategy (frequency, parameters, and sample distribution).
Justification of sampling sites	Information available from field workers and laboratory staff verbally	Need to revise sampling site selection and their justifications, with potential additions or removal of sites, as necessary.
Sampling strategy for each site (how sampling is conducted, where it is conducted and what is sampled)	The information can be made available through documentation transmitted from CPCB to the laboratories	Sampling strategy for each site should be improved and tailored to its specific conditions and objectives.
Sampling methodology (global & site-specific)	The global methodology is available through documentation provided by CPCB to the laboratories. However, the site-specific methodology appears to be unavailable, except for the designated water quality parameters.	While field workers appeared to know where, what, and how to sample at each site, it is unclear how this information is documented. They carried no paperwork to validate their site-specific approach. Fieldwork protocols are too general, lacking site specificity, and are challenging to apply in practice.
Sampling routes/logistics		Sampling site locations are documented; the route is not documented. Additionally, this information does not appear to be integrated into a database.
Workflows between planners, field workers, laboratories, data users	Not clear where & whether the information is documented	Information not available
Communication plan within the WQMP		Information not available
Plan for communicating data outside of the members of the WQMP		Information not available
Data validation processes	Documented	See Table 4

Table 2: Managing field data (sampling contexts) of Ganga WQMP

Criteria	Status	Notes
Field notes	Yes (paper only & incomplete information)	**Included:** Sampling site name, geographic coordinates, field worker names, date and starting time, water temperature, pH, oxygen titration, general weather conditions, and general notes. **Not included:** Sampling end time, site-specific protocols, sampling objectives, sampling tools and their protocols, standardized weather information (sky type, precipitation, wind, ambient temperature, atmospheric pressure), probes used, probe calibration details, and photos.
Field notes management	Yes, in folders (paper only)	Field notes are not transcribed into any type of data management system, except for pH, water temperature and dissolved oxygen results, which are transcribed into the CPCB LIMS (see block 6)
Field notes use (laboratory)	Yes	Field notes are made available to laboratory personnel
Field notes use (data users)	Yes	Field notes are made available to scientists interpreting the data

2.3 Recommendations for quality-controlled Data Acquisition

Table 3: Recommendations for Data Acquisition

Recommendation	**Link to Methodology**
Data acquisition strategies (rationale of the WQMP) must be made transparent and linked to the sampling sites	**2.4.1** Understand and document design and objectives of the WQMP
Following information must be documented and integrated into the IFMLD: ⇒ Sampling objectives ⇒ Sampling site justification ⇒ Sampling frequency ⇒ Sampling recurrency ⇒ Site specific sampling strategy ⇒ Site specific sampling methodology ⇒ Sampling routes	**2.4.2** Integrate the information into an IFMLD
Sampling personnel must be trained according to site-specific sampling strategies and methodologies, and the laboratory should have a master sampling expert to ensure training, QC/QA of the fieldwork for all the staff members. ⇒ Ensure communication with the laboratory staff on precautions to take in the field such as sample conservation overfilling bottles, etc. ⇒ Ensure that sampling staff is equipped specifically to the site requirements (regarding safety, transportation, and sampling). ⇒ Ensure that every laboratory has a designated area for sampling equipment storage and maintenance. ⇒ Ensure that logistics are being respected, for example- ⇒ Ensure that field note data are transcribed and communicated correctly into an IFMLD.	**2.4.3** WQMP specific training of field and laboratory staff using the information within the IFMLD

2.4 Implementation Methodology

2.4.1 Understand the WQMP objectives and how to design them

Figure 2 presents the main questions which must be asked and documented to understand and document the design and objectives of an existing WQMP. This information provides the essential foundation on which all quality control and quality assessments can and must be based. Indeed, this information is essential to all managers and planners of the WQMP, the field staff, the laboratory staff and data analysts, the water experts (e.g., limnologists, ecologists, or hydrologists), and any potential data user.

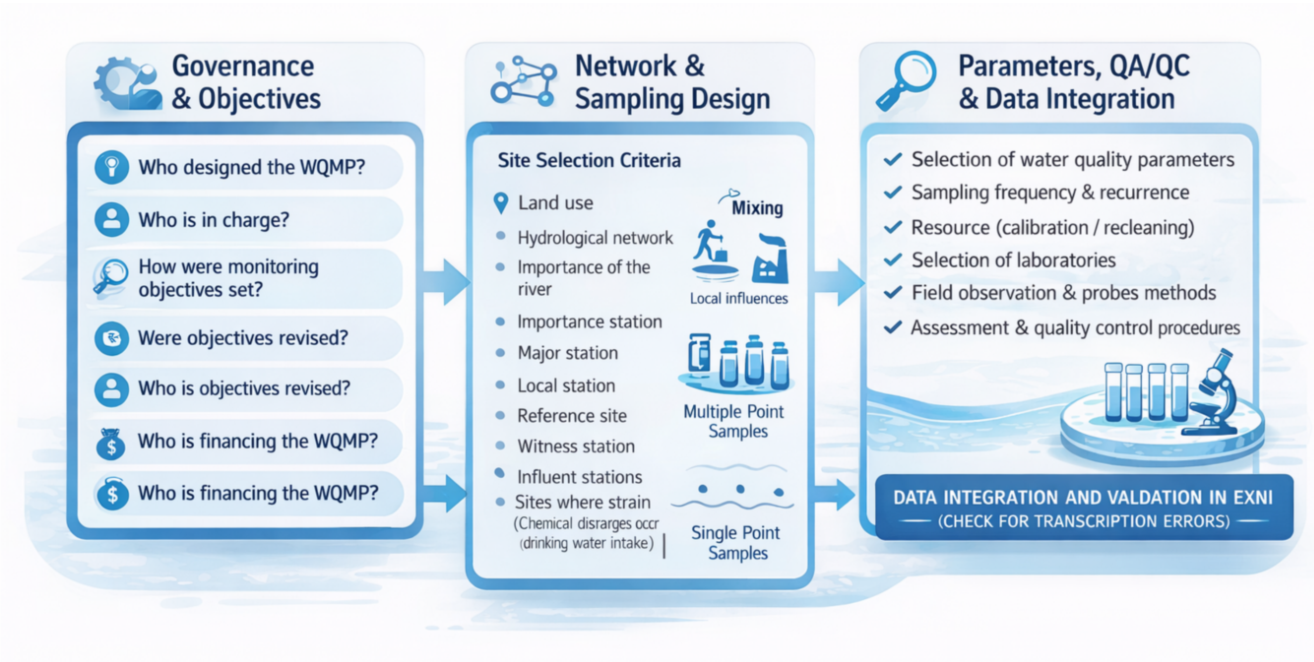


Figure 2: Questions to be asked (& integrated) into the system to understand the design of a WQMP (derived from Behmel et al. 2019)

2.4.2 Integrate the information into an IFMLD

The following figures illustrate how this information can be documented in an IFMLD.

⇒ **STEP 1:** It is important that the rationale, objectives and design of the WQMP are well-documented. This includes start date, description of the project (Water Monitoring Program e.g. NWMP, SWMP etc.), specific objectives, as well as any documents produced during the planning or operation of the WQMP. It is important that the data is acquired and managed according to the themes and future needs identified in this step (Figure 3A, 3B).

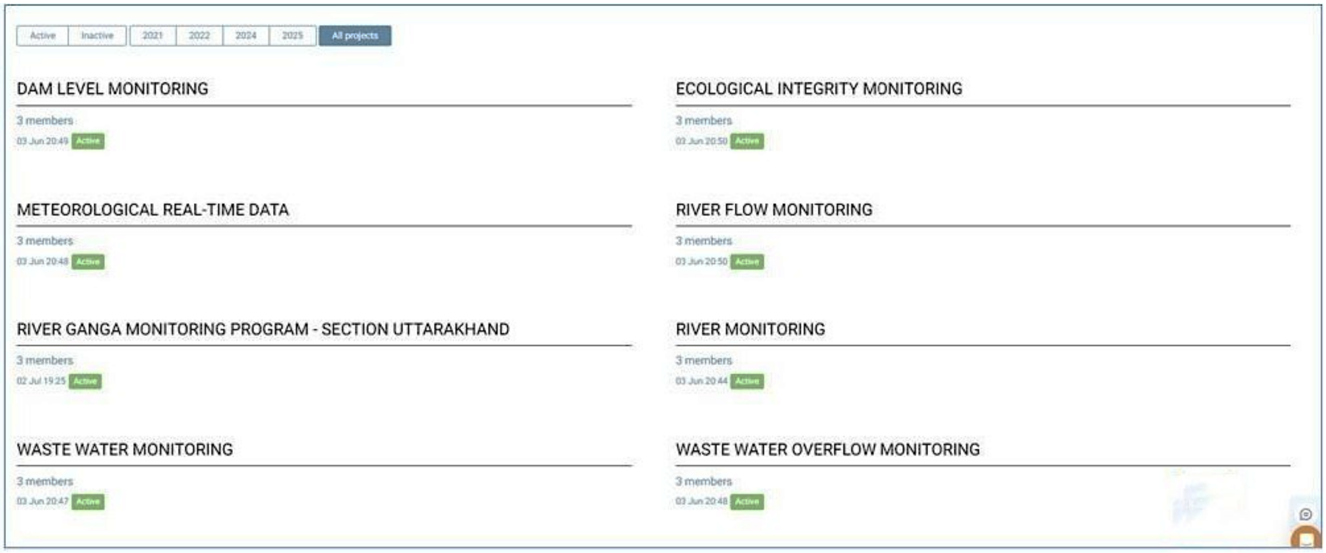


Figure 3: Project-based data management- Information to describe project *(sample screenshots from Enki, proprietary software from WaterShed Monitoring Inc., with permission. Other software providers with similar systems are Aquatics Informatics (WIMS) and Earthsoft (EQuIS)).*

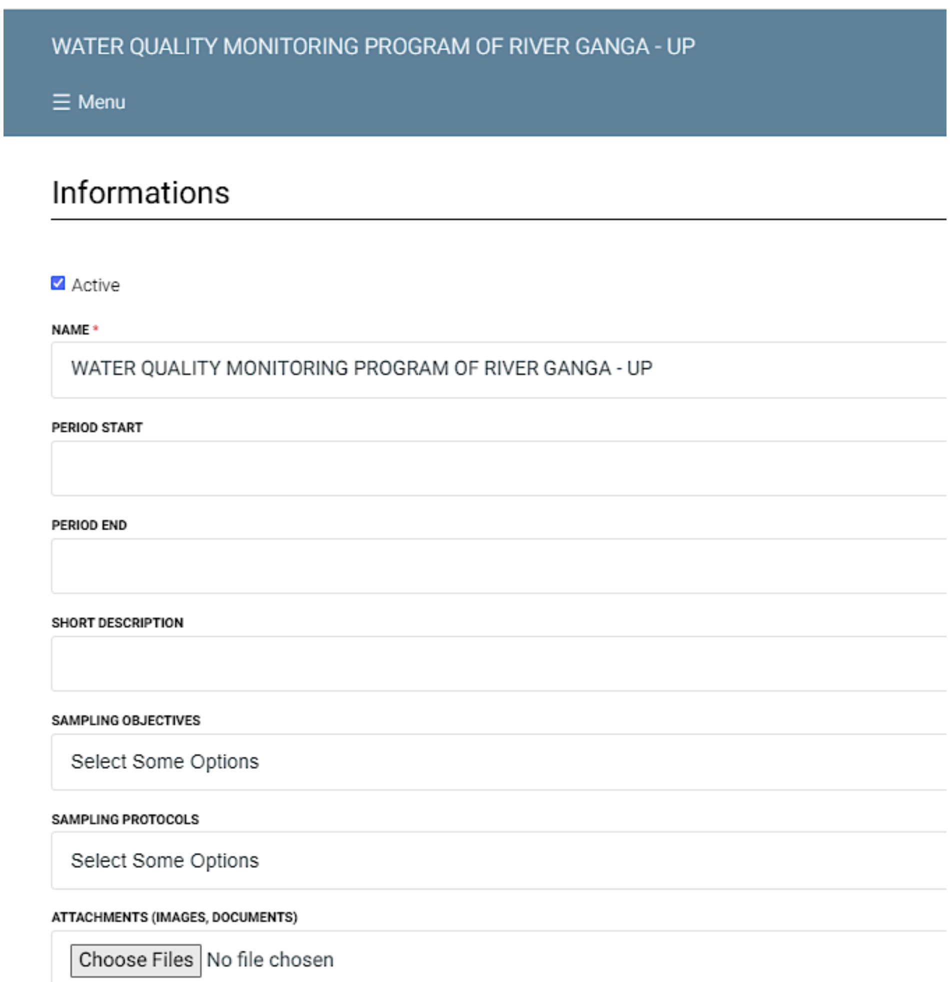


Figure 4: Dashboard of projects within a WQMP. Data from these projects can be used based on the user's access.

Optimizing the WQMP and analyzing the data it produces requires information on the watersheds within the territory of the WQMP, as well as the waterbodies that are being monitored. The list of information which can be made available can be very long, and should, in many cases, be managed by a Geographical Information System (GIS) (e.g. Figure 4). It would minimally include: Land use, geological information, topography, bathymetry, climatic regions, watersheds, and hydrological networks.

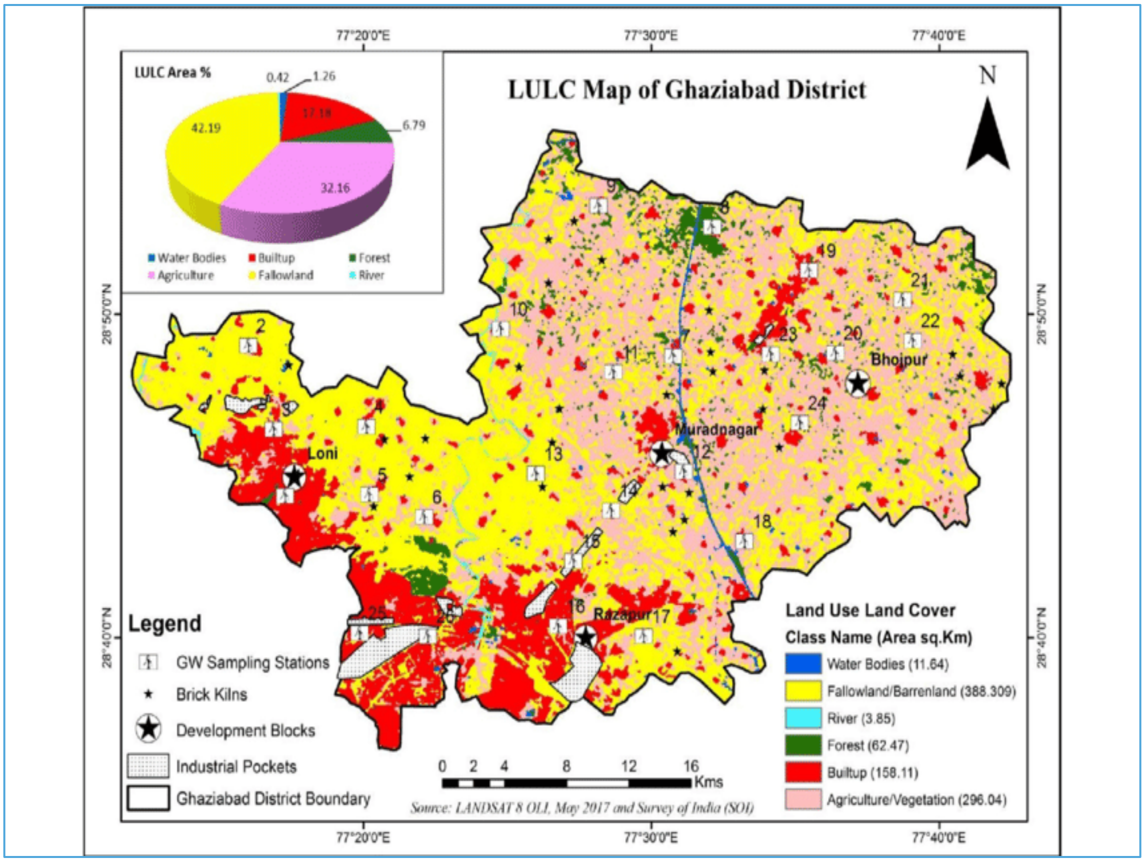
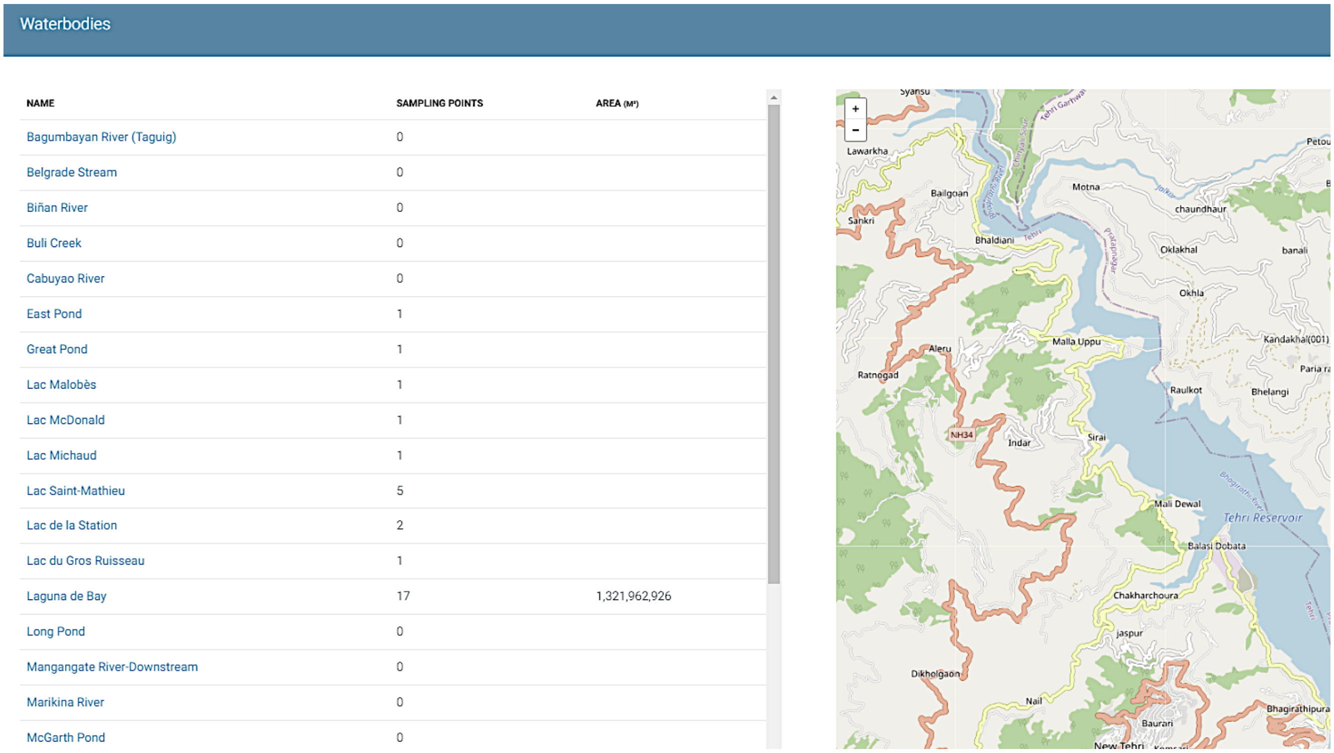


Figure 5: Land use land cover (LUCL) map of Ghaziabad District, Uttar Pradesh, including position of ground water sampling stations (Tyagi, S. et Sarma, K 2021).

Furthermore, the IFMLD should include the delimitation and description of watersheds and waterbodies with which sampling sites may be associated. More specifically, this includes- Name of the watershed, size of the watershed (km2), name of the waterbody, description of the waterbody, existing studies on the waterbody, bathymetry and other relevant information helpful to interpret the results of the WQMP (Figure 6).



Waterbodies

NAME	SAMPLING POINTS	AREA (M²)
Bagumbayan River (Taguig)	0	
Belgrade Stream	0	
Biñan River	0	
Buli Creek	0	
Cabuyao River	0	
East Pond	1	
Great Pond	1	
Lac Malobès	1	
Lac McDonald	1	
Lac Michaud	1	
Lac Saint-Mathieu	5	
Lac de la Station	2	
Lac du Gros Ruisseau	1	
Laguna de Bay	17	1,321,962,926
Long Pond	0	
Mangangate River-Downstream	0	
Marikina River	0	
McGarth Pond	0	

Figure 6: Minimum information on waterbodies and watersheds which must be integrated into the IFMLD

⇒ **STEP 2:** Identify sampling objectives and provide sampling site justifications.

It is essential to know what entity is responsible for sampling a specific area of the WQMP. In the WQMP of River Ganga described here, the sampling sites have been clearly assigned to laboratories under the local authorities. This information is necessary to be able to trace back the results from the WQMP, and to validate whether all information has been correctly reported in the IFMLD.

It is important to remember that monitoring objectives guide all the decisions on sampling site locations, selection of water quality parameters, sampling frequency and recurrence, and sampling site specific strategies. Moreover, sampling objectives provide a clear orientation for how the data can be analyzed, used and communicated. Therefore, it is essential that monitoring objectives are as specific as possible and are reported in the IFMLD for each sampling context. Data results which cannot be connected to an objective are unusable. Note that the WQMP described here provides relatively clear objectives (Figure 6). However, the results of the data should be reported along with the objectives.

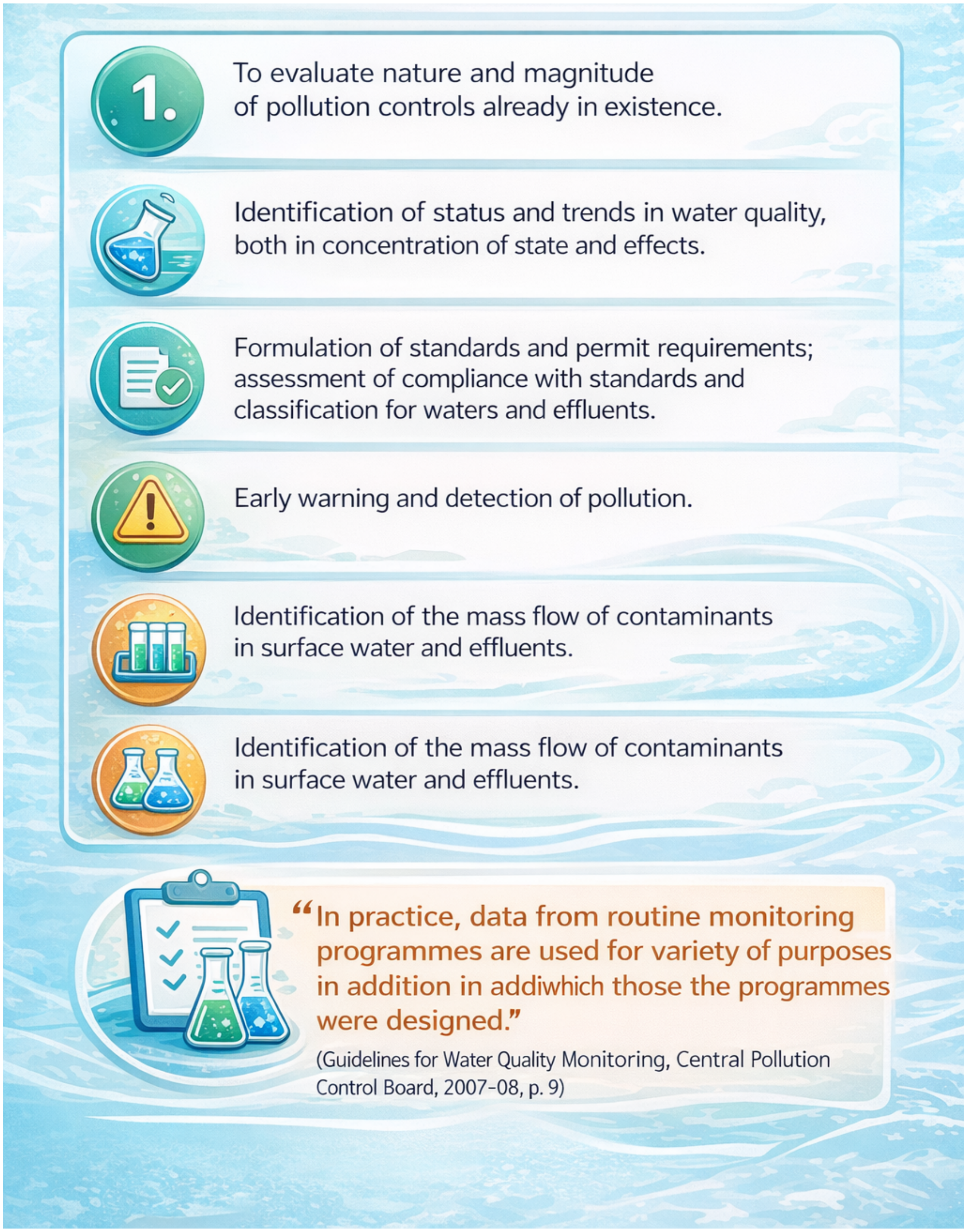


Figure 7: Current objectives of the River Ganga WQMP

Sampling site justification explains how a sampling site represents the specific objectives for future data analysis. A IFMLD can integrate all this information and link it to the sampling contexts (see next section).

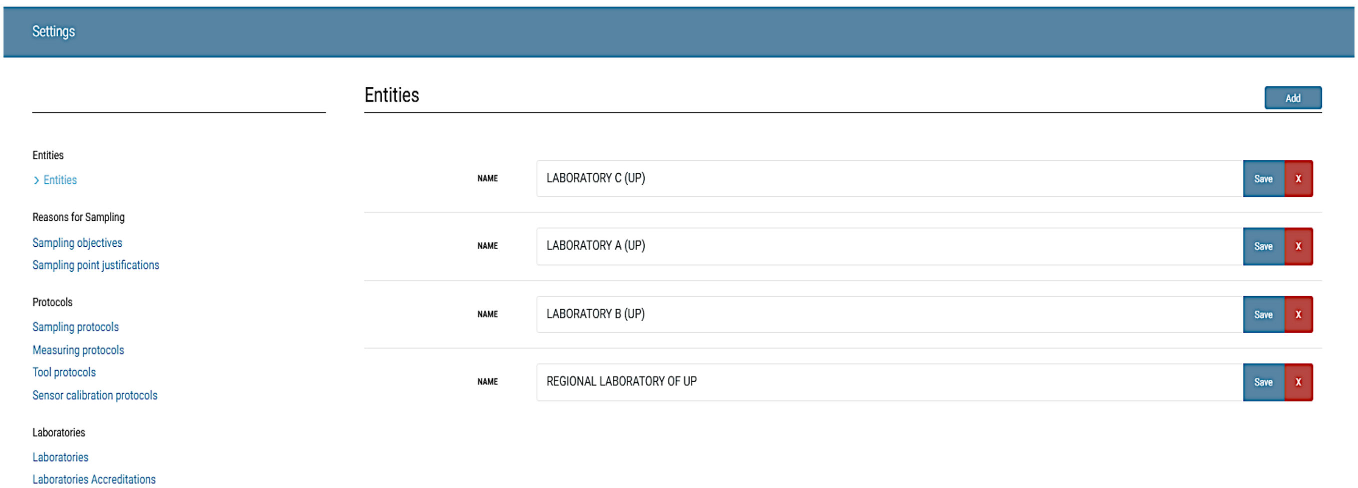


Figure 8: Entities responsible for sampling, sampling objectives and sampling site justifications must be integrated into the IFMLD.

In order to ensure that the sampling strategy is suitable for the objectives and the configuration of a sampling site, and that it is applied consistently to the same sampling site, the information illustrated in Figure 9 must be included in the IFMLD.

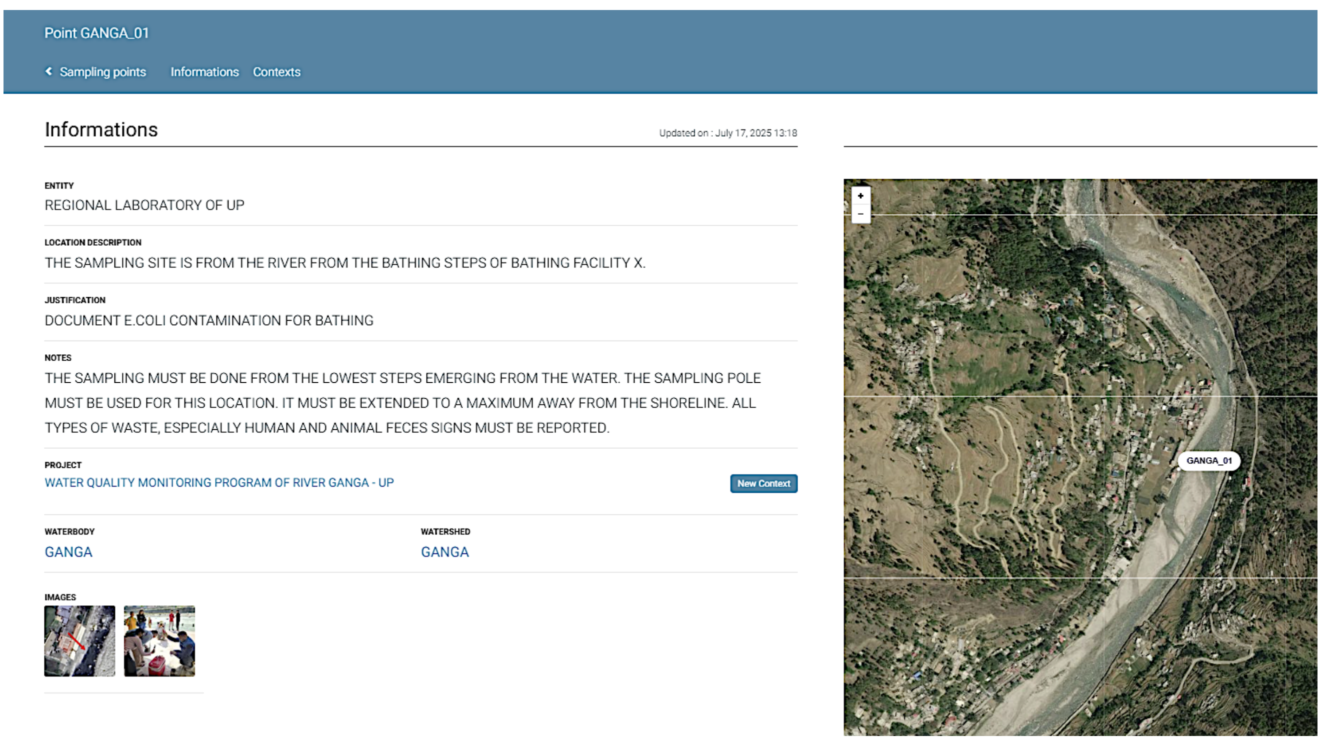


Figure 9: Information that needs to be integrated to fully describe a sampling site and site-specific sampling strategies.

⇒ **STEP 3:** Integrate information on protocols for sampling, tools, sensor calibration, and maintenance into the IFMLD (Figure 10).

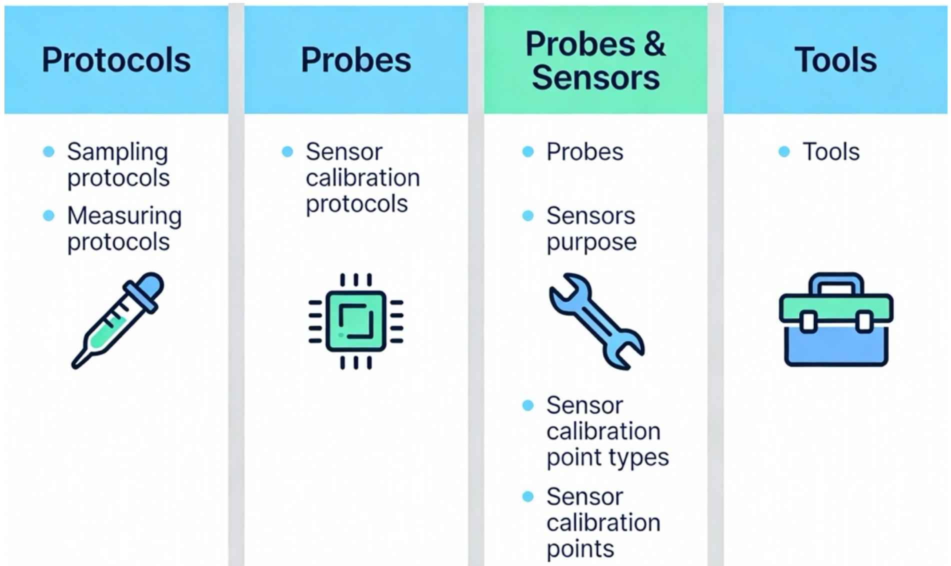


Figure 10: Information on protocols for sampling, tools, and sensors, as documented in the IFMLD

It is of fundamental importance that all the systems contributing data to the IFMLD (laboratories, probes, sensors, etc.) are parameterized in accordance with the parametrization of the IFMLD to ensure data integrity and to avoid ambiguities, particularly for (1) Parameter Names, and (2) Units of Measurement. That is to say, the IFMLD should provide the parameter names, and all the parameter names from the LIMS and other data sources should be parametrized the same way. If that is not possible, a system of corresponding parameter names (allowed aliases) can be set up. A description of the parameter name is also important and will also help data users. Finally, each parameter and unit of measurement must have a unique identifier, which may be drawn from the SI codes, or a national code .

⇒ **STEP 4:** Once all systems have been parametrized, it is possible to generate sampling routes and sampling documentation for the field workers. This documentation also helps the laboratory to prepare for sampling (e.g., providing bottles and materials, coolers to ensure temperature control).

The information which should be included is:

- Sampling sites to be visited, along with all the above-mentioned information about the sampling site.
- The sampling method to be employed.
- The sampling tools to be used.
- The parameters to be sampled.
- A picture/description of the site.
- Safety considerations at the sampling site.

- Route planning/description from laboratory to sampling sites, and between sites.

2.4.3 WQMP specific training of field and laboratory staff using the information within the IFMLD

It is essential to provide the field and laboratory staff with information on all the sampling logistics, including sampling strategy, equipment to be used, sampling bottles to be provided, etc. This is necessary to ensure that the field and laboratory staff are fully aware of:

- ✓ WHAT to bring to the field
- ✓ WHY the sampling is done
- ✓ WHERE to sample
- ✓ HOW to sample (remember to always be site-specific!.
- ✓ HOW to ensure cooling and safe transportation of samples (Figure 13) *(It is essential to bring back the samples in a timely manner; time-frames are parameter-specific and should be discussed with the laboratory; ensure the cool boxes are clean and cooled to between 4ºC and 10ºC).*
- ✓ HOW to transport sampling materials (Figure 14).

Note that every single site will have specific challenges, and that conditions may change due to changes in water level, or in human activities upstream or around the sampling site. Therefore, it is essential that the organization planning the WQMP does a pre-sampling tour with the field staff to identify best sampling strategies, and what information should be shared with the planners if any changes occur. Any alternative sampling strategies to be used should be clearly explained, as well as how to document changes to sampling. It is extremely helpful to have an IFMLD that allows the inclusion of sketches and pictures to show how the sampling site is configured under different water level conditions, and that provides information about what to do in these different conditions. Such an IFMLD also allows the field staff to report on how things were done, and why, including the possibility to add photos. Figure 11 below illustrates a well-documented sampling context.

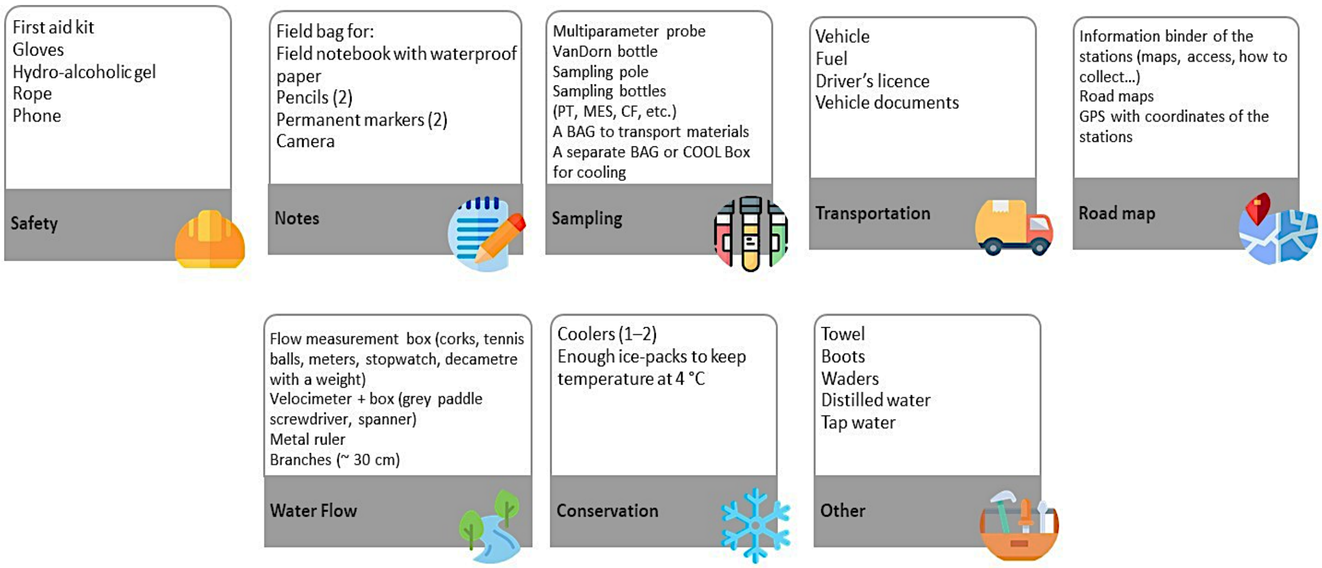


Figure 11: Checklist of WHAT to bring to the fieldwork.

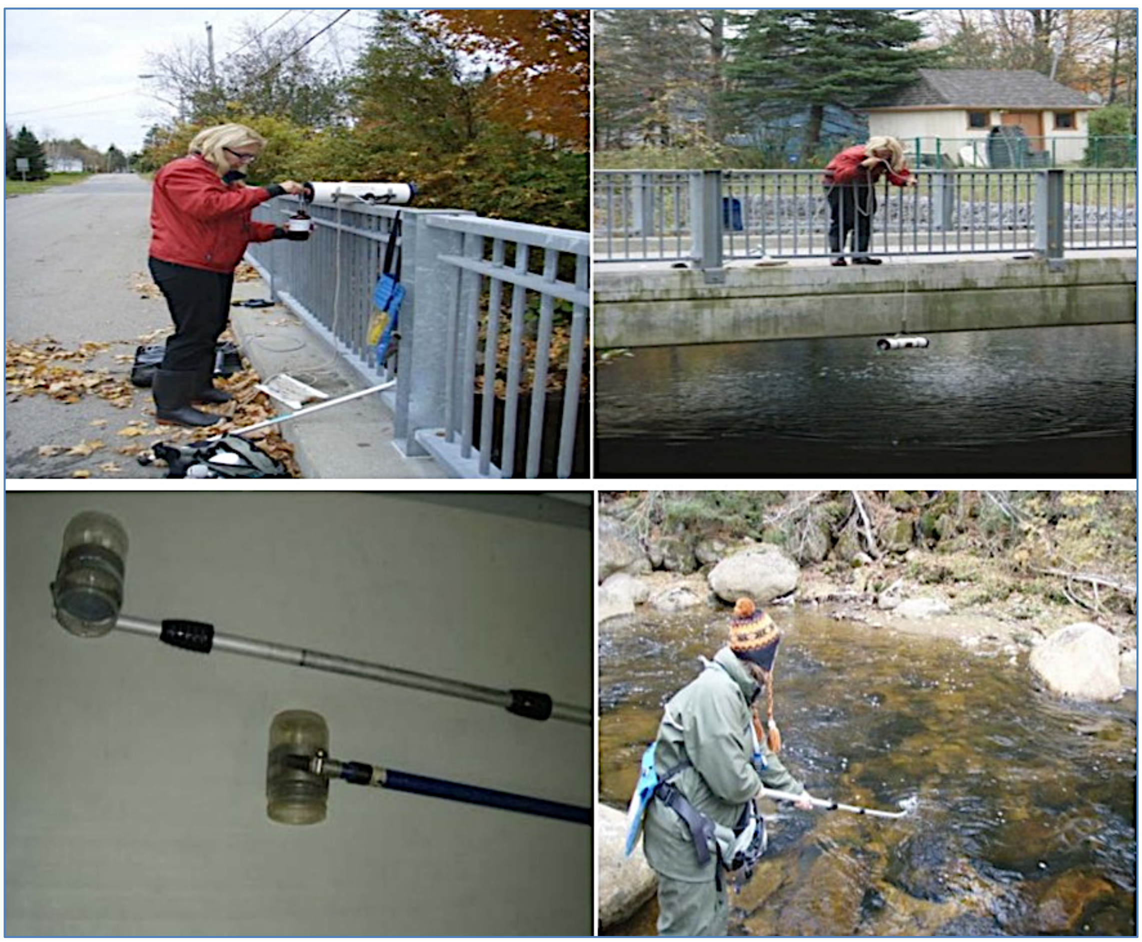


Figure 12: How to sample is site-specific. Example above: Sampling from a bridge with a Horizontal Alpha Bottle. Example below: Sampling from the shoreline with a home-made extendable sampler.

Important!

A cool box is opened only to introduce ice bags and filled sampling bottles

A cool box should not be too big, nor too small. It should be adapted to the number of bottles and optimal number of ice-packs to keep it cool without freezing the samples.

Packing instructions can be included. Good packing will also prevent breaking of bottles.

A cool box is not designed to transport sampling materials

Figure 13: How to ensure optimal cooling conditions and safe transportation of sample bottles. Do not transport sampling tools and materials, personal items, or food or beverages in the cool box.

Figure 14: Example of a well-packed sampling toolbox. A backpack is also very acceptable, especially for river sampling. The box or backpack must be kept clean, organized, and be exclusively used for sampling materials. It should not be used to carry personal items, or food and beverages.

3. Methodology for quality control of data and meta-data management practices within WQMP

3.1 Introduction

Data management within a WQMP must include:

- SELECTION OF THE OPTIMAL DATA MANAGEMENT SOFTWARE
- PARAMETRIZATION OF THE DATA MANAGEMENT SOFTWARE
- DATA VALIDATION PROCEDURES AND PROCESSES
- DATA INTEGRATION PROCEDURES AND PROCESSES
- USER ACCESS MANAGEMENT
- EXCHANGE OF DATA BETWEEN THE LABORATORIES (LIMS), PROBES (FIELD NOTES/ HANDHELD/ CLOUD-BASED SYSTEMS), OR OTHER SOURCES, AND THE DATA MANAGEMENT SOFTWARE

The following methodology will provide:

⇒ Key information for selecting the optimal data management software (key functionalities).

⇒ Key elements of the parametrization need. These will be limited, as parametrization always depends on the software used; however, there are standards which should be common to all systems used to manage data and metadata within a WQMP.

⇒ Best practices and workflow for data validation procedures and processes. We will not present an exhaustive review of potential statistical or deep learning techniques for data validation.

⇒ Data integration procedures and processes to be used by field workers, laboratory data integrators, data validators, and data contributors. The focus will be on manual and batch imports, as well as some automated 'red flags' for errors/omissions which should always be included in the software.

⇒ User access management. While this is always software-specific, we will present some of the main considerations.

3.2 Current situation of the Ganga WQMP

Currently, the Ganga WQMP does not have an IFLMD.

Individual laboratories do not have a LIMS to manage and organize data related to the laboratory processes. Instead, they integrate their testing results into a centralized LIMS which does not account for differences in how individual laboratories document processes. Since the main purpose of a LIMS is to streamline and optimize various laboratory processes, the lack of a LIMS means that laboratories are unable to digitally document the processes that lead to the results they share. *(Refer to Appendix B for more details on the current situation.)*

3.3 Recommendations for quality- controlled Data Management

Table 4: Recommendations for Data Management in WQMP

Recommendation	Link to Methodology
Every laboratory must be equipped with its own laboratory information system (LIMS).	**3.4.1** LIMS
The organization responsible for the WQMP must be equipped with an IFLMD (see previous section for more details).	**3.4.2** IFMLD
Results from the laboratories must be transferred through unified electronic transcripts / application programming interface (API) to the IFLMD	**3.4.3** Data transfer

3.4 Implementation Methodology

3.4.1 LIMS

Each laboratory must use a LIMS with information about:

- Laboratory process management from the sampling to the production of the laboratory analysis report

- Laboratory instrument maintenance
- Human resource management, including basic training, training on analytical techniques, and follow-up training
- Documentation of quality control and quality assessment, and follow-ups.

In other words, the LIMS will enable a laboratory to digitize all their operations, which will improve efficiency, contribute to producing defendable results, facilitate accreditation and auditing processes, and facilitate data sharing with partners, authorities and clients.

A LIMS should include the following elements:

1) Easy and swift automation of processes to minimize errors and increase efficiency and output.
    a) As a minimum (but not exhaustive) list, the LIMS should:
    b) Verify related parameters to avoid errors in analytical reports (e.g., conductivity vs ion analysis, BOD vs COD, total nitrogen vs ammoniacal nitrogen).
    c) Make it impossible for an analyst to perform an analysis if he/she isn't adequately trained, by impeding either the operation, or the entering of results.
    d) Make it impossible to perform data analysis if the validation of the analytical method isn't completed or it is erroneous, by requiring that necessary pre-conditions are completed.
    e) Allow for automatic transfer of data from analytical equipment (balances, spectrophotometers, chromatographs, ICP, etc.) to avoid transcription and other errors.
2) Full compliance with regulations and ISO Norms, such as ISO 17025.
3) Automated sample tracking and inventory management.
4) Data access restricted to authorized users, and a complete read-only audit trail. Moreover, restrictions should be related to analyst or supervisor status (analysts have access to methods for which they are adequately trained, or for which they have the appropriate level of security). Other controls can be applied (prioritization of samples close to their expiration date, etc.) to assure quality of analysis.
5) Calibration and maintenance schedule for ready-to-use instrument list. Controls can also be added to avoid use of non-functional equipment.
6) Improved measures for quality control and quality assurance (QC/QA). A series of analyses must include the quality control samples (number and kinds) provided

for the quality control procedure. In addition, these measures must reach the thresholds required in the procedure for the batch of analyses to be validated.

7) Digitized signatures to enhance data security. Only authorized persons can sign the reports in a LIMS. These authorizations are related to the methods (e.g., a microbiologist cannot sign a chemistry report, the person signing must have adequate knowledge of the methods included in the report, etc.). The LIMS should be able to consider staff absences to forward the report to be signed to another signatory.
8) Automated conversion of raw data and population of the calculated results into the certificate of analysis.
9) Application programming interface (API) to an IFMLD (E.g. Enki, WIMS, EQuIS).

3.4.2 IFMLD

Ideally, a single IFMLD should be made available for each WQMP. In the case of the River Ganga WQMP, an IFMLD for each state could serve the purpose of improving data management of the WQMP. An important consideration is that an IFMLD centralizes all environment-related water monitoring data, including the rationale of the WQMP, documentation, field work, sampling contexts, laboratory results, laboratory tracking (including information from the LIMS), and metadata. Such a system avoids transcription errors in environmental data, and includes communication features, procedures for streamlined data use, etc. It also allows for data publication and sharing under specific conditions and protocols. It therefore contributes to producing and storing defendable, comparable and shareable data.

With these purposes in mind, the IFMLD should be parametrized which means setting it up with the following information:

- ⇒ Rationale of the WQMP (see section 2.4 Methodology).
- ⇒ Description of sampling sites with site-specific sampling methodology
- ⇒ Description of waterbodies and watersheds
- ⇒ Protocols for field measurements, instrument use, and maintenance and maintenance follow-up
- ⇒ Protocols for use and maintenance of field sampling tools.
- ⇒ Harmonization of parameters, units of measurement, and unique identifiers with those in the LIMS.

A simplified workflow showing the main steps for onboarding an IFMLD is illustrated in Figure 15. The main elements of such a system are presented in Table 5: Essential components of an IFMLD. Additional information on onboarding is IFMLD (and LIMS) specific and can be provided by the software provider.

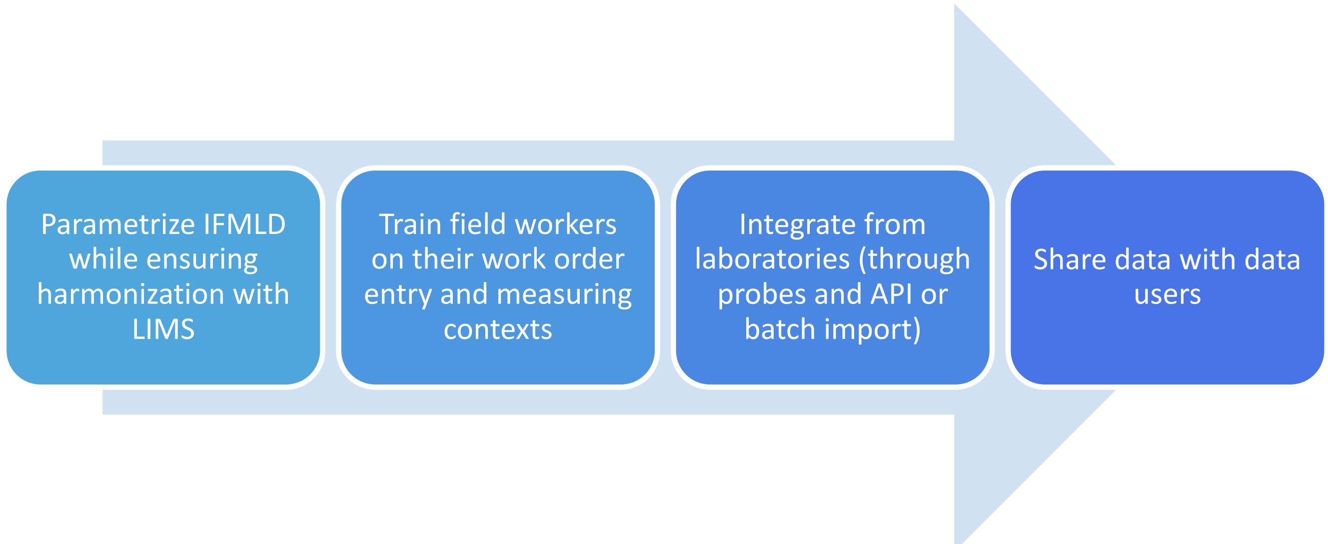


Figure 15: Simplified workflow of parametrization and use of the IFMLD.

Table 5: Essential components of an IFMLD.

Components	Reason for including the component
Data import ▪ Entry of measurement contexts ▪ Probe data import (probe types to be specified) ▪ Lab data import ▪ Manual data entry (entry of measurement contexts and field measurements) ▪ Dashboard (global activity monitoring)	This feature allows the centralization of information from the field work sheets, and the results from the field instruments (e.g., temperature) and field tests (e.g., oxygen titration) and links this information to the rationale of the WQMP (e.g., sampling objectives), field testing methodologies, and results from the laboratories.

Data entry into the parametrization tables
▪ Descriptive data
▪ Measurement contexts
▪ Detailed probe information
▪ Detailed laboratory information
▪ Methodologies
▪ Metadata
▪ Documents in PDF, Excel or JPG format
▪ Additional descriptive information (e.g., details about the location of the sampling station, observations when taking measurements, etc.)
▪ Geographical data
    ▪ sampling points
    ▪ waterbodies/distribution network
    ▪ watersheds/distribution network
	This feature enables all the members of the WQMP to provide and obtain as much metadata as possible, and to document what is done, how, where, why, with what tools, and by whom.

Data visualization and retrieval ▪ Wide range of filters available ▪ Export in CSV, JSON and XML formats ▪ Visualization of data in graphical form ▪ Sharing data in a controlled manner with (1) a database with the same structure for raw data and metadata, and (2) a GIS system / Website to communicate results (Please refer to data sharing protocol methodology).	This feature: – enables the user to retrieve the data from a source with validated and standardized data and metadata. – enables the specialists in a WQMP (e.g., limnologists) to interpret the data correctly. – ensures that the data do not need to be manually transcribed from a PDF format, thus reducing transcription errors. – ensures that the data user is aware of how results below detection/quantification levels are being treated (e.g., divided by two, eliminated). – ensures that the statistical tools are calibrated with the same algorithms, thus achieving the same results. – enables the user to save queries, thus making the process repeatable for faster reporting.

Management of access levels
▪ Possible user profiles:
    ▪ *Administrator* (with full rights)
    ▪ *Contributor* (with the right to add measurement contexts)
    ▪ *Visitor* (with or without retrieval rights)
	Ensures controlled integration and sharing of the data.

Languages ▪ Language preference can be easily changed (in one click)	Choice of interface in English, Hindi, or other languages can be helpful for data integration and exploitation.
User training	
Technical Support ▪ Online, directly in the app, through a support ticket system	Ensures that the software vendor's involvement does not end with the sale but includes subject matter expertise on water quality monitoring and technical support.

3.4.3 Data transfer

Currently, each laboratory in the Ganga WQMP manually integrates all data into the central LIMS and other data systems (please refer to Appendix B).

Once the LIMS and IFMLD have been put into place, the next step to be considered is to allow data from the LIMS to be automatically imported to the IFMLD by implementing an application programming interface (API), a set of protocols that enables different software to communicate and transfer data. Note that it is also possible to extract full data sets from the LIMS and batch import them into the IFMLD.

This is particularly easy when the systems have standardized parametrization (i.e., same parameter names, units of measurement, and sampling site names). The chief objective here is to avoid, as much as possible, manual data entry of laboratory results from one database into another. An example is provided in Appendix A.

The data grouped together in the IFMLD can then be shared with other systems, or data users can be given controlled access. Figure 16 provides an overview of a suggested implementation of LIMS and IFMLD systems with a suggested data flow.

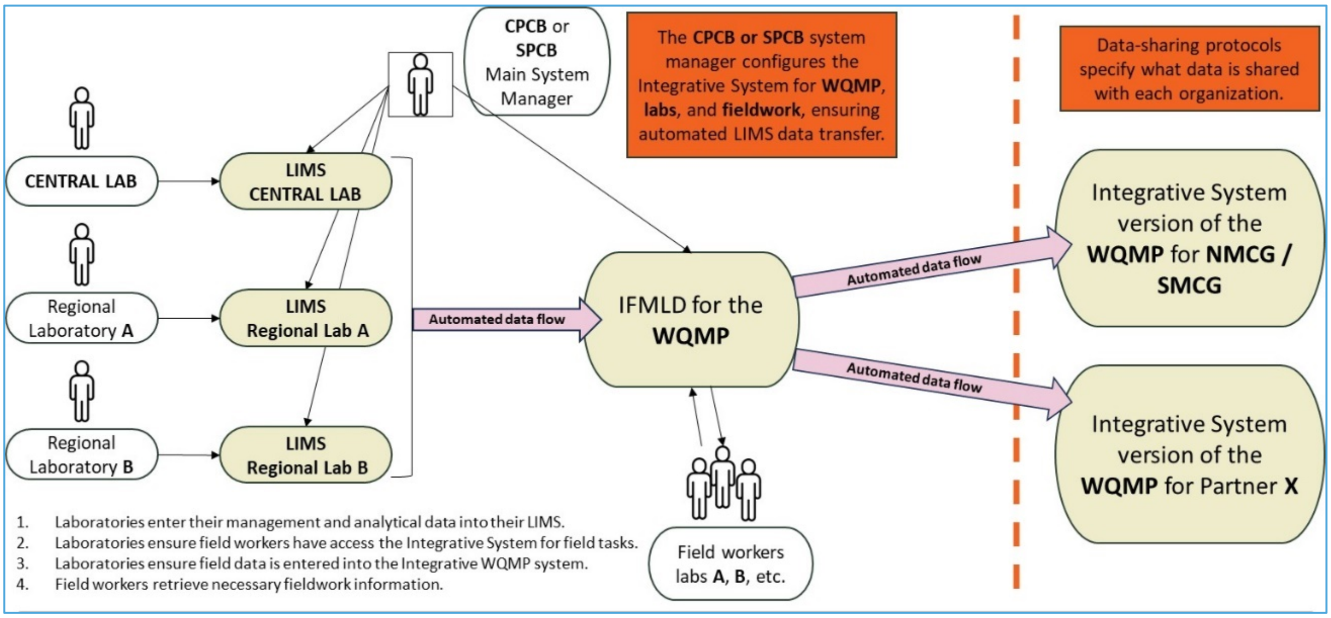


Figure 16: Suggested approach for data management software used by different actors and laboratories in the WQMP.

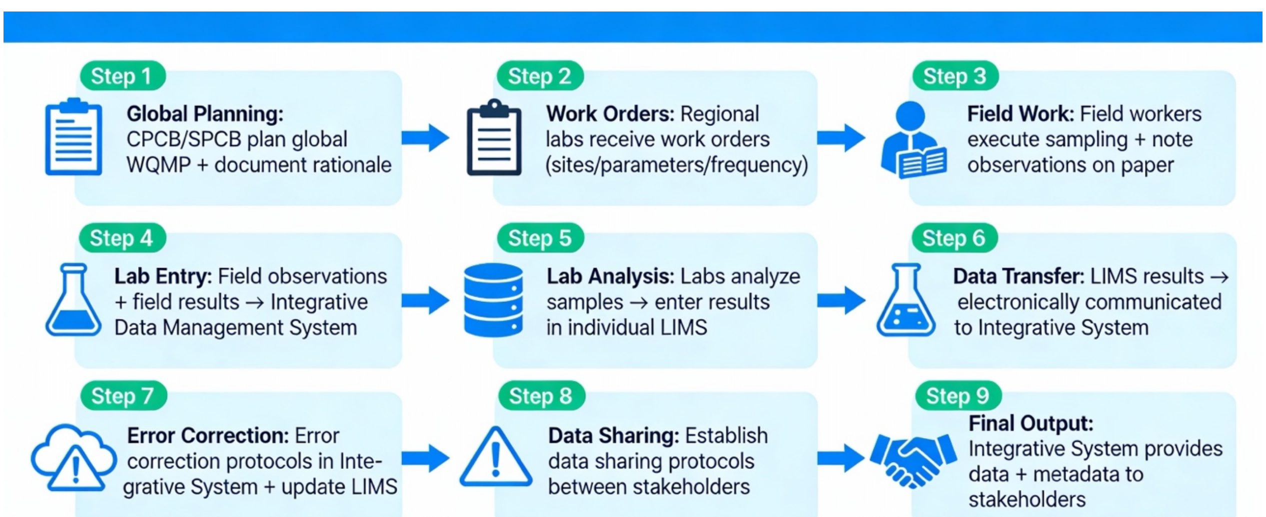


Figure 17: Suggested workflow description

4. Methodology for quality control and processes for data validation and analysis within WQMP

4.1 Introduction

Normally, the data validation process is completed before data analysis. However, the data validation and data analysis steps are closely related, and should not be fully separated, since inconsistencies in the data sets may come up during data analysis that cannot always be explained through the sampling context. Thus, the workflow must include a step for going back to the source of the data (e.g., probes or laboratories), to double-check potential errors.

In the proposed methodology, we will define the different levels of data validation, and the main processes to follow. We do not include data analysis methodologies (e.g., how to apply a given statistic), but we will include some key elements that should be taken into account when selecting a data analysis method which is coherent with the original objectives of the WQMP and the data set used for a given analysis.

4.2 Current situation of the Ganga WQMP

Currently, in the Ganga WQMP, the only data validation procedure in place is between the Regional and Central Laboratories. The Regional Laboratories manually integrate the analytical results into the LIMS of the Central Laboratories, while manually integrating them in parallel into the EWQDES from CPCB, and potentially other portals. Should a question arise about the results, the data correction follows the procedure described below:

To the best of our knowledge, the field note data is only used by the laboratories for analytical purposes and is not communicated to the data users. Probe / Sensor data from the field (e.g., pH and temperature) are manually entered into the LIMS from the Central Laboratories, without information on maintenance and field conditions. Also, the current procedure for correcting data anomalies takes place via emails between the Central and Regional laboratories and occurs only 30 to 45 days after sampling was completed. When the data users receive the processed data report, it does not include metadata.

In addition to this, the data user is not made aware, and does not have access to, any information about the rationale of the WQMP.

The minimal information the data user needs to have is:

a. Sampling objectives
b. Sampling site justification
c. Geographical context of the sampling site
d. Land use of the watershed feeding the waterbody
e. Sampling strategy for the site
f. Meteorological conditions
g. Date and time of sampling
h. The expected statistics to be applied (according to objectives and time-series)
i. The thresholds / indicators to be used for analysis.

Note that the data sharing protocols are discussed in Section 5.

4.3 Recommendations for Data Validation and Analysis

If the recommendations for the LIMS and IFMLD described in Section 3 have been followed, they will lay the foundations for data validation. Therefore, the recommendations in this section focus on

⇒ Data validation processes during the data integration phase for the IFMLD.
⇒ Data validation processes during the data analysis phase.

Table 6: Recommendation for Data Validation within WQMP

Recommendation	Link to action in the methodology
Each laboratory executing field work much put in place a process of ensuring timely and correct recording of sampling contexts into the IFMLD.	**4.4.1** Process for recording sampling contexts
The IFMLD must have a data entry log that shows potential errors when batch-importing, with a process in place to implement corrections.	**4.4.2** Process for data integration validation
A data validation process should be implemented for the data analysis phase.	**4.4.3** Data validation process for the data analysis phase

4.4 Implementation Methodology

4.4.1 Process for recording sampling contexts

The process is illustrated in Figure 17.

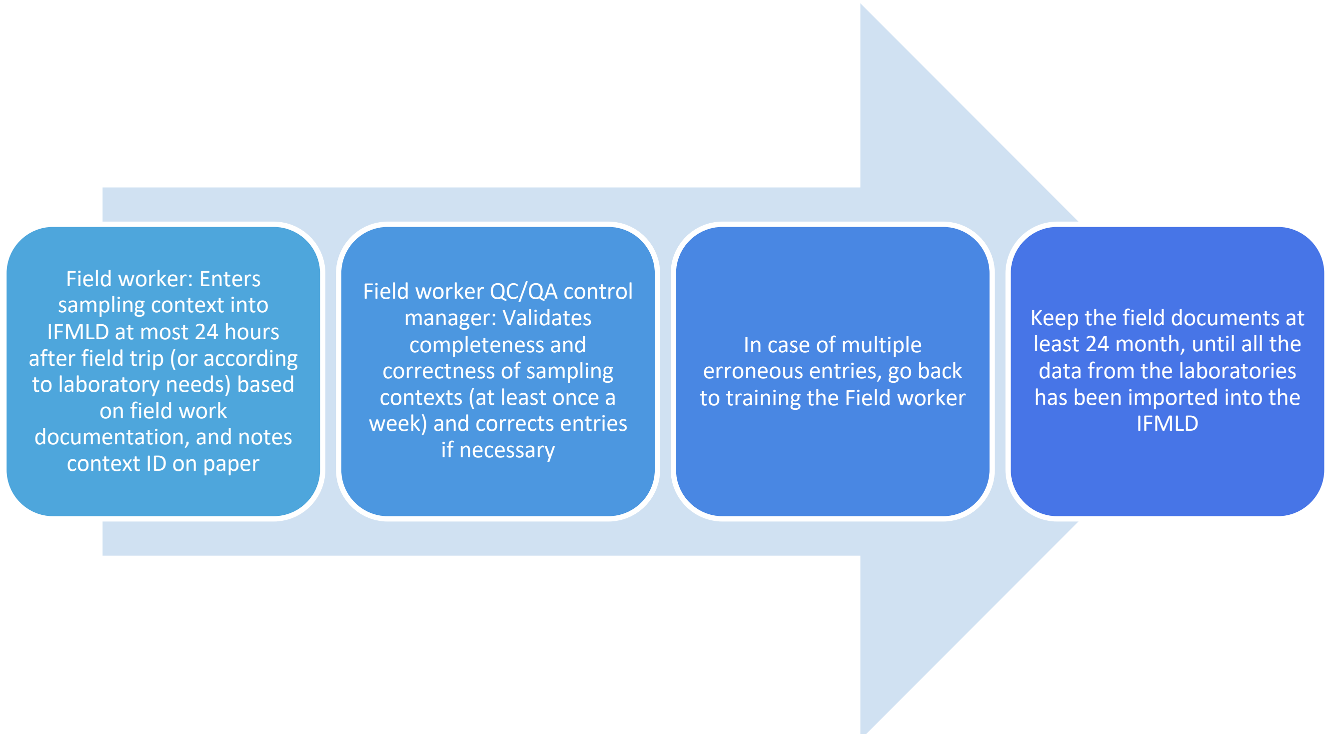


Figure 18: Workflow for recording and validating sampling context.

4.4.2 Process for validation of data integration

Manual data entry of laboratory results into any type of system other than the laboratory's LIMS should always be avoided. Data transcription is always prone to transcription errors. Therefore, we recommend that results from the LIMS be imported by an API or by batch import into the IFMLD.

The key to integrating the results from a manual water quality monitoring program is complete and correct sampling contexts. The data from the laboratory will be integrated into the sampling context using several essential connectors:

- ✓ Sampling site name (which must match exactly between the LIMS and the IFMLD).
- ✓ Sampling date (which is not the same as the date of reception of the sample at the laboratory, or the date the analysis was done in the laboratory).
- ✓ If multiple samplings occurred on the same date, same site, the LIMS must also contain the information on the beginning and end of sampling from the sampling context.

Furthermore, it is essential that the parameter names and units of measurement delivered by the LIMS match the parameter names and units of measurement in the IFMLD. Normally, if both systems have been parametrized based on international standards, this should not be a problem. Mismatches in the above-mentioned elements are possible. If they occur, the steps described in the table in Appendix C must be followed in order to ensure data integrity.

4.4.3 Data validation process for the data analysis phase

The data user (e.g., limnologist of the WQMP, statistician) must be made aware and given access to the entire context of the data provided. Pure statisticians without any knowledge about water quality data in the natural environment should be made aware that environmental data can have several outliers which should be considered as results rather than anomalies. It is precisely these anomalies which must be validated during the data analysis phase[1].

We present three examples to illustrate the data validation process.

Example 1:

The data analyst applies a principal component analysis (PCA) to a given set of data from the Saint-Charles River watershed between 2011 and 2013. The objective of the WQMP was to obtain the spatio-temporal variability of water quality under dry and wet conditions. The results in Figures 19 and 20 show several extreme outliers. As a starting point, Figure 20 provides some contextual information on the sampling sites implicated. By positioning these sites on a map, along with sampling site justification, the analyst may uncover some elements to help with data interpretation.

Still, the question remains: Do these extremes represent errors in the data (occurring in the laboratory or during the sampling), or can they be attributed to other factors (e.g., conditions in the watershed, meteorological events, human activities)?

---

[1] Information from CPCB:
CPCB holds review meetings with all regional offices of SPCBs/PCCs regarding submission of data under NWMP w.r.t data completeness, delay & related quality assurance aspects and provide its minutes of meeting and status report to respective Regional Directorate and CPCB head office.

CPCB regularly reviewed and analysed NWMP data. In case of any outlier or abnormal value, the concerned agencies were informed through various communications (e.g., dated 02.06.2024, 06.04.2021, etc.) to rectify data gaps and ensure the uploading of correct data on the portal.

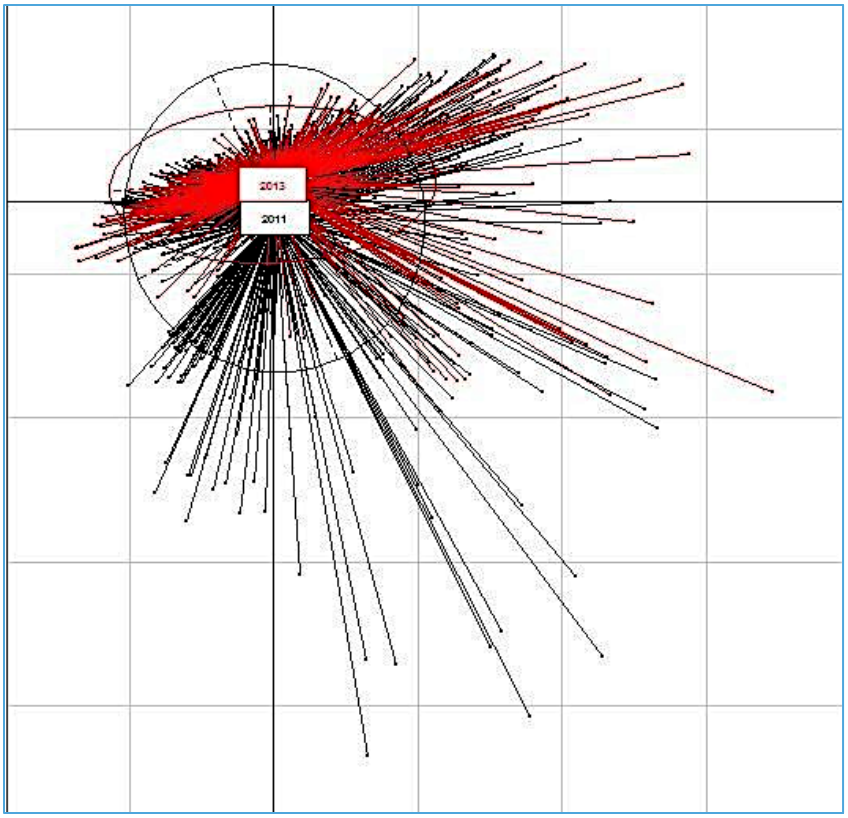


Figure 19: PCA of water quality data from the Saint-Charles River watershed between 2011 and 2013.

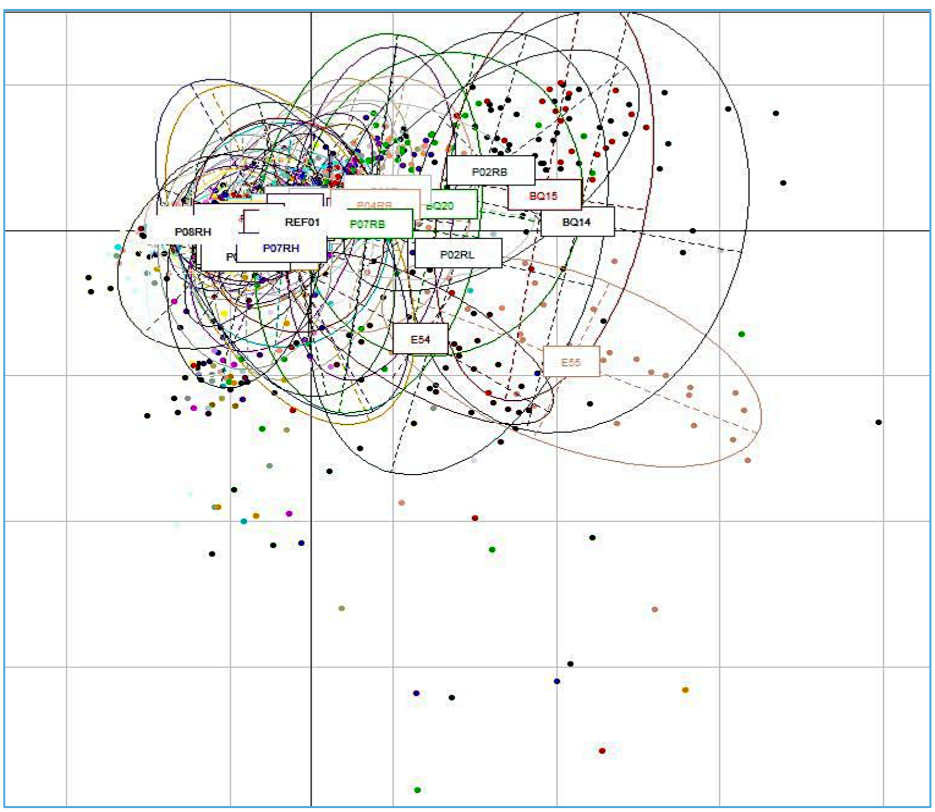


Figure 20: PCA of water quality data from the Saint-Charles River watershed between 2011 and 2013, including names of sampling sites.

The validation steps to be followed are:

1) *Can the extremes be explained by the conditions of the sampling sites within the watershed?*
    a. *If yes, proceed to the next question.*
    b. *If no, go back to validating the sampling context.*
2) *Can the extremes be explained by the meteorological conditions during the sampling?*
    a. *If yes, obtain the information on weather events to provide context for the results.*
    b. *If no, go back to validating the sampling context.*
3) *Can the extremes be explained by human activities?*
    a. *If yes, obtain the information on the human activities to provide context for the results.*
    b. *If no, go back to validating the sampling context.*

If none of the contextual information can explain an extreme result, the result should be:

1) *Signalled to field worker's supervisor to validate whether there was contamination of the sample due to bad sampling methods (e.g., using a bucket with a cord for total phosphorus or metals, putting the hand into the sampling device).*
2) *Signalled to the responsible person at the laboratory who must then proceed to an investigation on the laboratory side and potentially cancel the result and update both the LIMS and the IFMLD.*

Example 2:

A data analyst creates a box-plot chart directly in the IFMLD (Figure 21). There are three examples of extreme outliers for stations E01, E02 and E10. To verify the extreme outliers of E01 and E02, the analyst follows same procedure described in Example 2. However, the outlier of E10 (specific conductivity at 0 µS/cm) represents an IMPOSSIBLE result. It is therefore fully erroneous, and must be flagged, and the source of the error must be traced. It may result either from a probe which was used incorrectly, or from a laboratory result in distilled water.

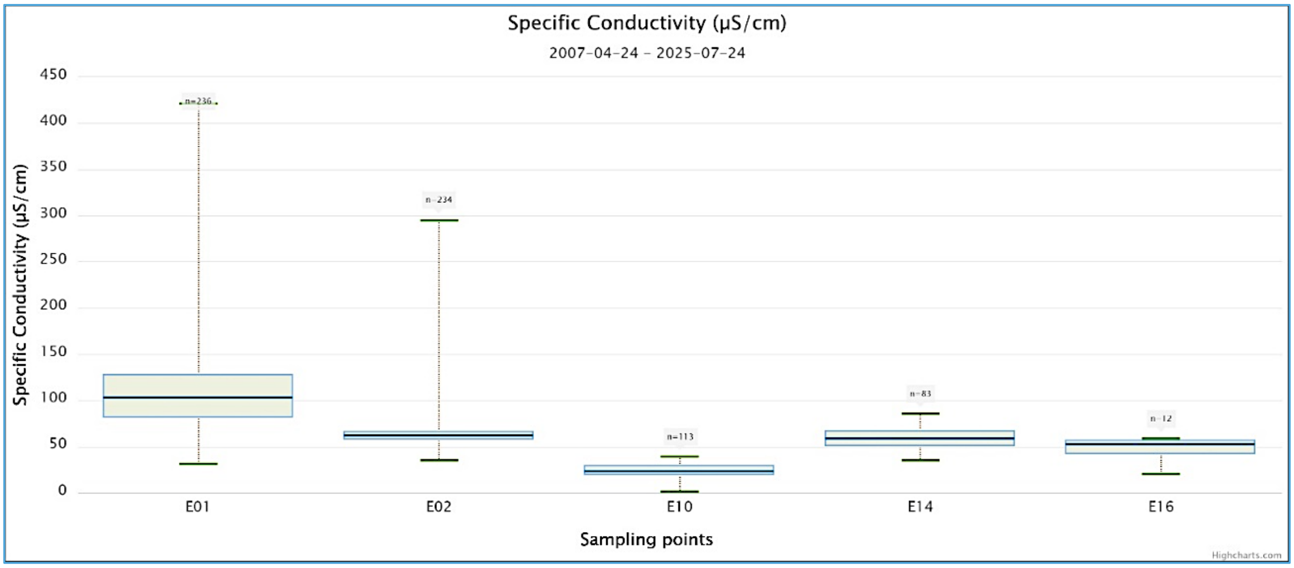


Figure 21: Results from five stations in the Saint-Charles River watershed, with three outliers at stations E01, E02 and E10.

Example 3:

The data analyst creates a box-plot chart directly in the IFMLD for measurement of E. coli at four stations located upstream to downstream on the same river, E08, E07, E55, and E01 (note logarithmic scale). Large outliers occur for all stations. This type of outlier is very typical for E. coli measurements and is not necessarily a sign of problems with the sampling or with the laboratory analysis.

To verify the outliers, it is essential to understand the geographical location and the justification of each sampling site. E08, E07 and E01 are sampling sites on the river that experience high dilution effects, meaning that it is not always possible to capture all E. coli during sampling, especially if mixing is not optimal, which is usually the case.

E55 is the effluent of a wastewater treatment plant. Extremes outliers, ranging from near 0 CFU/100 mL to 30 000 CFU/100 mL can be explained by having information about the Date and Time of sampling, and knowledge about the set-up and operations of the wastewater treatment plant which is being sampled. In this case, low concentrations in E. coli could be explained if the population living in this area works downstream from the houses connected to the wastewater treatment plant and are not at home at the time of sampling, thus increasing the efficiency of the system. Alternatively, the treatment plant may use UV lamps on some dates and not others. High concentrations in E. coli could be explained if the population is staying at home, if many tourists are contributing to wastewater charges (due to a holiday), if no UV lamps are used, of if there are difficulties in treating the wastewater correctly due to overload.

E08, E07 and E01 are river stations. Although their outliers are less extreme, the results are still highly variable. Knowledge of Date and Time of sampling and weather conditions are essential to interpret the results. For instance, higher E. coli levels are expected after heavy rains, compared to days following a dry spell. Because station E01 is downstream of the wastewater treatment plant, information on what is going on at E55 is particularly important to understand the results (especially if the sampling is timed in such a way that a correlation can be established between upstream and downstream results.

This example illustrates the necessity of knowing everything about the potential elements that may influence the results before questioning the sampling personnel or the laboratory.

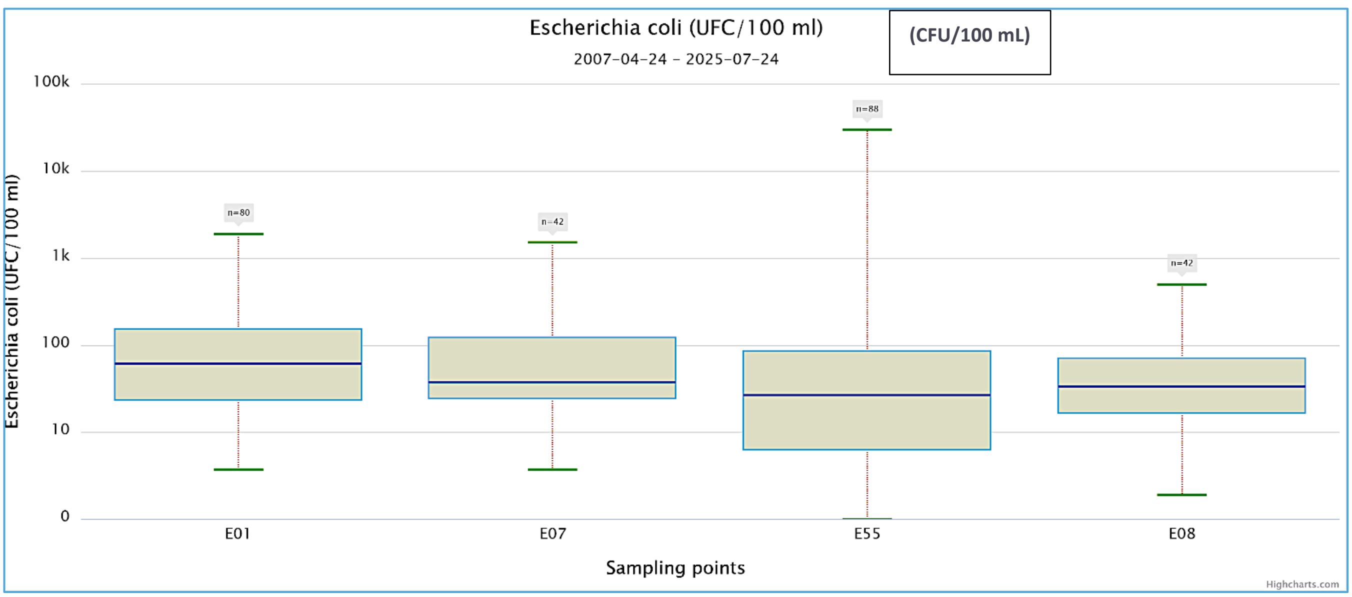


Figure 22: Example of results of Eschericia coli in CFU/100 mL along a river (upstream to downstream).

The overall validation and correction process should follow the steps described above and illustrated in Figures 23 and 24.

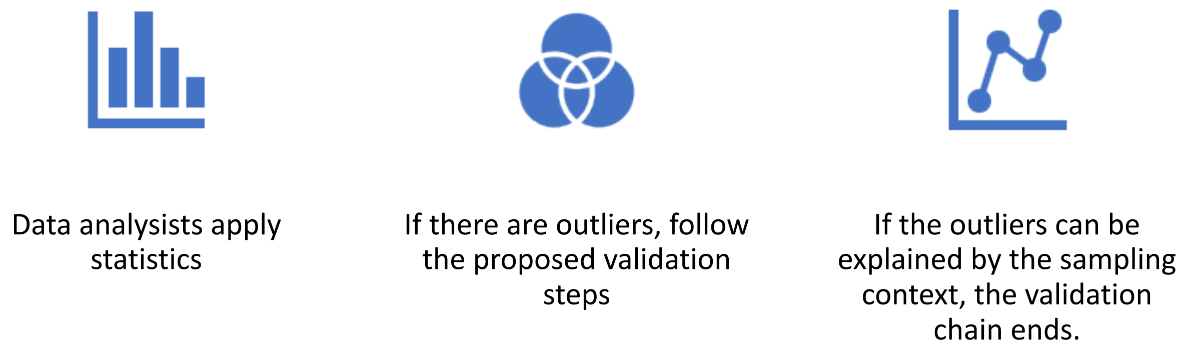


Figure 23: Data validation process 1.

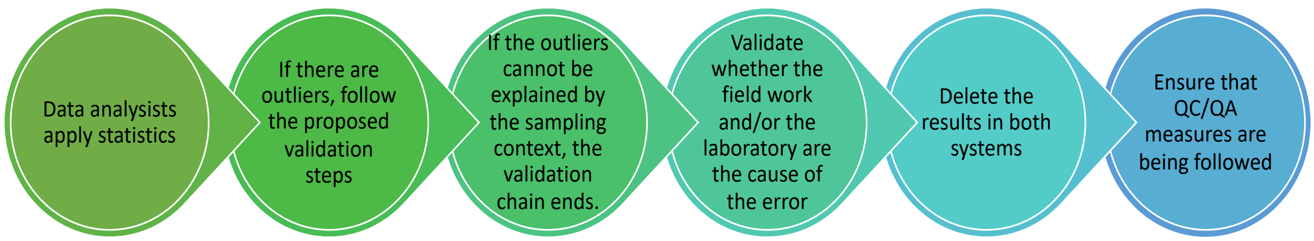


Figure 24: Data validation process 2.

5. Methodology for data sharing within WQMP

5.1 Introduction

Two categories of sharing of WQMP data should be considered:

I. Sharing the data with professional stakeholders, who know how to interpret and use data.
II. Sharing the data with non-professional stakeholders, with the assumption that they don't know how to interpret and use data.

Data sharing protocols should therefore always include a stakeholder analysis. This specific methodology will focus on data sharing protocols with professional partners of a WQMP, with a very specific focus on the WQMP and stakeholders under the NMCG-PTB project.

5.2 Current situation

The current situation is described in Table 6. Please refer to Appendix B for more information on the context of this table.

Table 7: Current situation of data transfer between CPCB, SPCB and NMCG/SMCG as of July 2025.

Criteria / Necessary information to be reported	State	Notes
Data transfer protocol	No	There is no clear data transfer protocol between CPCB and NMCG. Currently, CPCB provides data to NMCG upon request through PDF laboratory reports, which require manual transcription by NMCG staff, increasing the risk of errors. While CPCB can also provide data in Excel format upon request, the specific contents of these files remain unclear. Additionally, CPCB publishes some results on its website, but the data is limited, and no metadata is included. This also applies to SMCG Uttar Pradesh and UPPCB. SPCB transfer data to CPCB through EWQDES portal*.
Unified electronic format including metadata	No	No unified electronic format for data & metadata transfer has been established.
Access to the central database through a specific (limited) user access	No	Since no central database exists (including field data, etc.) access does not exist. In addition, NMCG does not have any type of access to the EWQDES of CPCB.
Data use protocol	No	There appears to be no established data use protocol. Key aspects remain undefined, such as how data should be used, whether results below detection or quantification limits should be reported, which statistical tools should be applied (as different tools yield varying results for the same data), and the limitations of the data in relation to the original objectives.

5.3 Recommendations for quality- controlled Data Sharing

Table 8: Recommendations for Data Sharing

Recommendation	**Link to the action in the methodology**
A data dissemination process must be established between the concerned parties	**5.4** Methodology

5.4 Implementation Methodology

The data dissemination process must include:

- ✓ Identification of the parties and persons which can and must receive information.
- ✓ A data use agreement between these parties and persons.
- ✓ An automated process of data sharing between these parties and persons.

Here, we will concentrate only on the basic methodology, for which the following pre-conditions are established and must be fulfilled:

PRE-CONDITION 1:
THE DATA AND METADATA ARE AVAILABLE THROUGH AN IFMLD.

PRE-CONDITION 2:
THE DATA USER IS A PROFESSIONAL WITHIN THE WQMP; HOWEVER, THIS PERSON MAY BE FROM AN ORGANIZATION DIFFERENT FROM THE ONE DOING THE FIELD WORK, LABORATORY WORK AND DATA INTEGRATION.

PRE-CONDITION 3:
THE DATA USER HAS ACCESS TO THE DATA AND THE METADATA.

PRE-CONDITION 4:
THE DATA USER HAS SIGNED DATA USE CONDITIONS (AN EXAMPLE IS PROVIDED IN APPENDIX C).

PRE-CONDITION 5:
THE DATA USER HAS BEEN MADE AWARE OF THE DATA VALIDATION PROCESSES (SECTION 4.4.3).

PRE-CONDITION 6:
THE DATA USER IS A VISITOR WITH EXTRACTION ACCESS TO THE IFMLD FOR THE PROJECTS IN QUESTION (FIGURE 24).

If all these conditions are fulfilled, the data user can then have access to the IFMLD's Observation Report website, where the data can be extracted and/or treated (Figure 25).

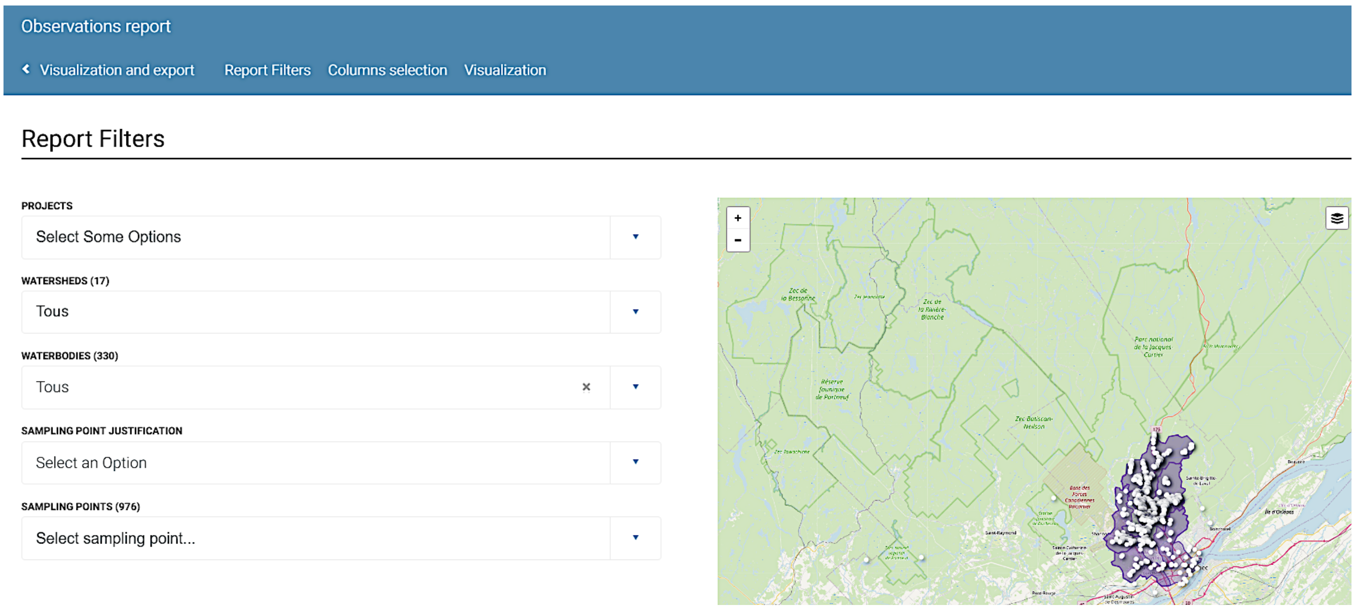
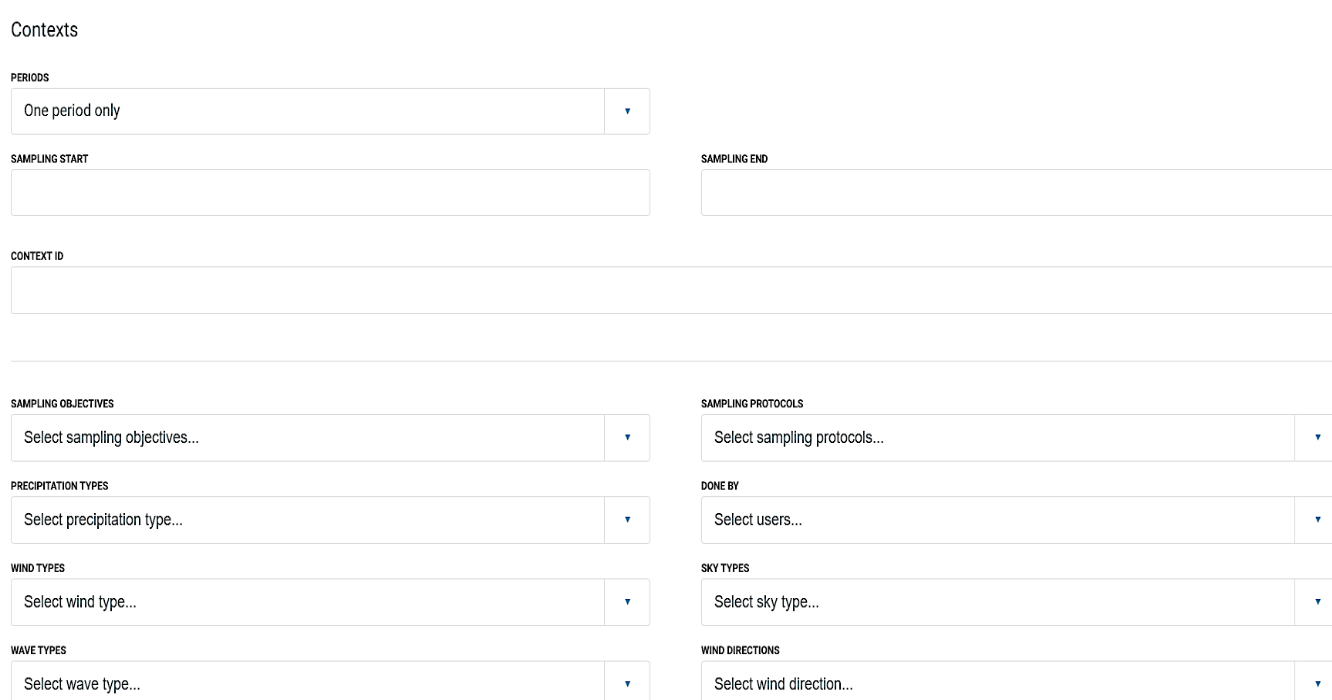


Figure 25: Filters to extract data and metadata.



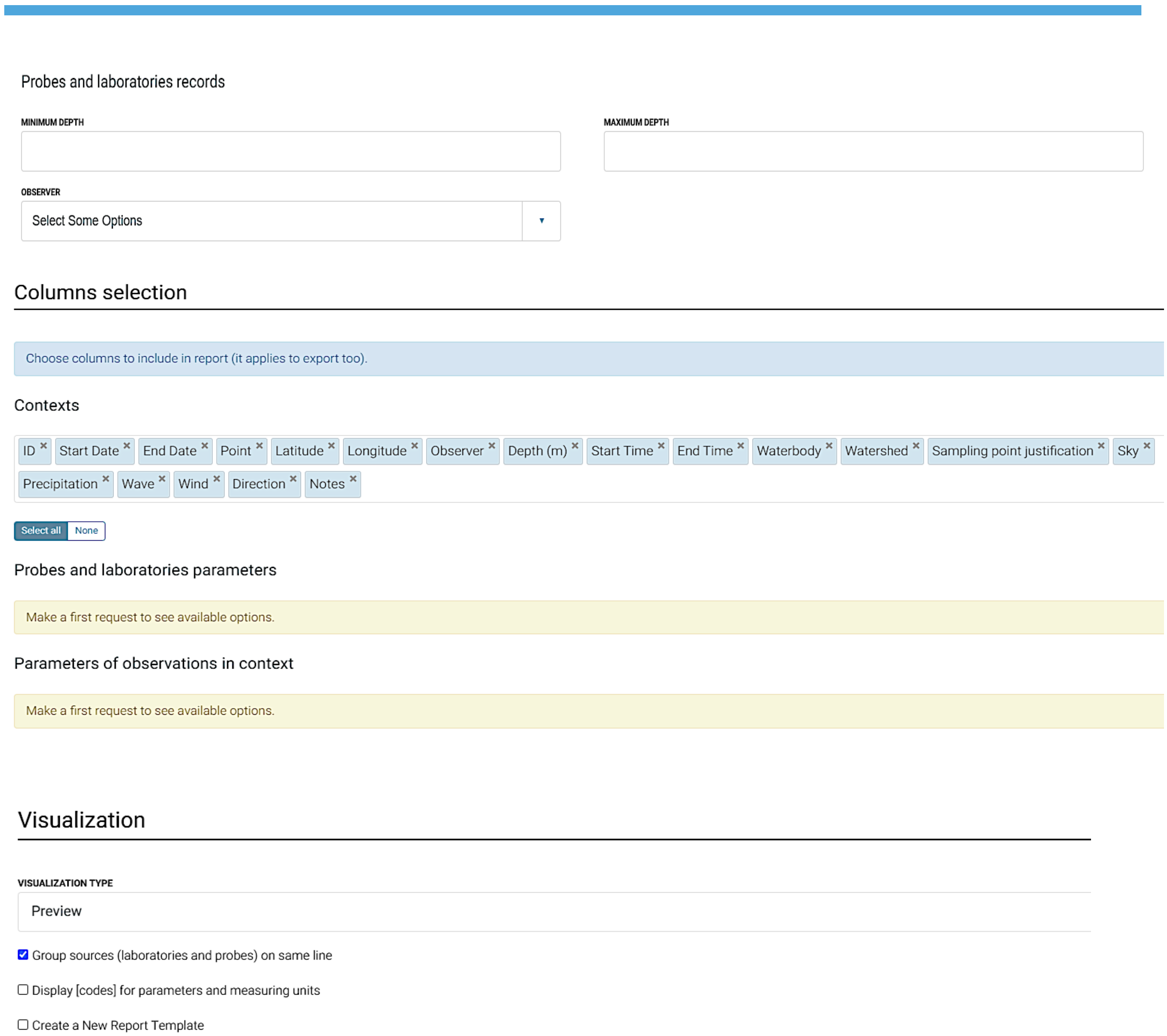


Figure 26: Example of an observation report website for data extraction and visualization by the data user.

6. References

- Behmel, S. 2010. “Proposition d’un programme de suivi de la qualité d’eau à l’échelle du bassin versant de la rivière Saint Charles, Québec.” Master’s Thesis, Québec: Université Laval. https://www.agiro.org/wp-content/uploads/Behmel2010_Programme_suivi_BV_saint_charles-1.pdf.

- Behmel, S., Damour, M, Ludwig, R, et Rodriguez, M. 2019. “Optimization of River and Lake Monitoring Programs Using a Participative Approach and an Intelligent Decision Support System.” Applied Science 9 (19): 4157. http://dx.doi.org/10.3390/app9194157.

- Behmel, S., M. Damour, R. Ludwig, et M. J. Rodriguez. 2016. « Water quality monitoring strategies — A review and future perspectives ». Science of The Total Environment 571 (novembre) : 1312-29. https://doi.org/10.1016/j.scitotenv.2016.06.235.

- Tyagi, S. et Sarma, K. 2021. « Seasonal variability, index modeling and spatiotemporal profiling of groundwater usability in semi-urban region of western Uttar Pradesh, India ». 80 (761). Environmental Earth Sciences. https://www.researchgate.net/figure/Land-use-land-cover-LULC-map-of-Ghaziabad-district_fig3_355859469.

7. Appendices

Appendix A:

Example of a data sharing protocol: Good Practice Guidelines for Data Management Policy- World Water Data Initiative, Bureau of Meteorology, Government of Australia (**Good practice guidelines – water data management policy**)

Appendix B:

Overview

Preliminary report based on:

- Four workshops with regional laboratories in 2024
- One workshop with a benchmarking laboratory in Québec, Canada in 2024
- Two laboratory visits in Lucknow / One visit in Dehradun in 2024
- One workshop at National Mission of Clean Ganga in 2024
- One meeting with the former Director & Head of the Quality Council of India, Alok JAIN in 2024
- Field work observations in 2023 & 2024
- Exchanges during workshops held in 2023 & 2024

Prepared by:

Dr. Sonja BEHMEL, President & Founder of WaterShed Monitoring Europe

Disclaimer:

Data management for surface water quality monitoring in the Ganga Basin WQMP requires further information and validation from involved parties.

Important note:

This analysis only includes manual sampling for surface water quality data from UP and UK

Date of current report:
September 12, 2025

I. Context

II. Methodology

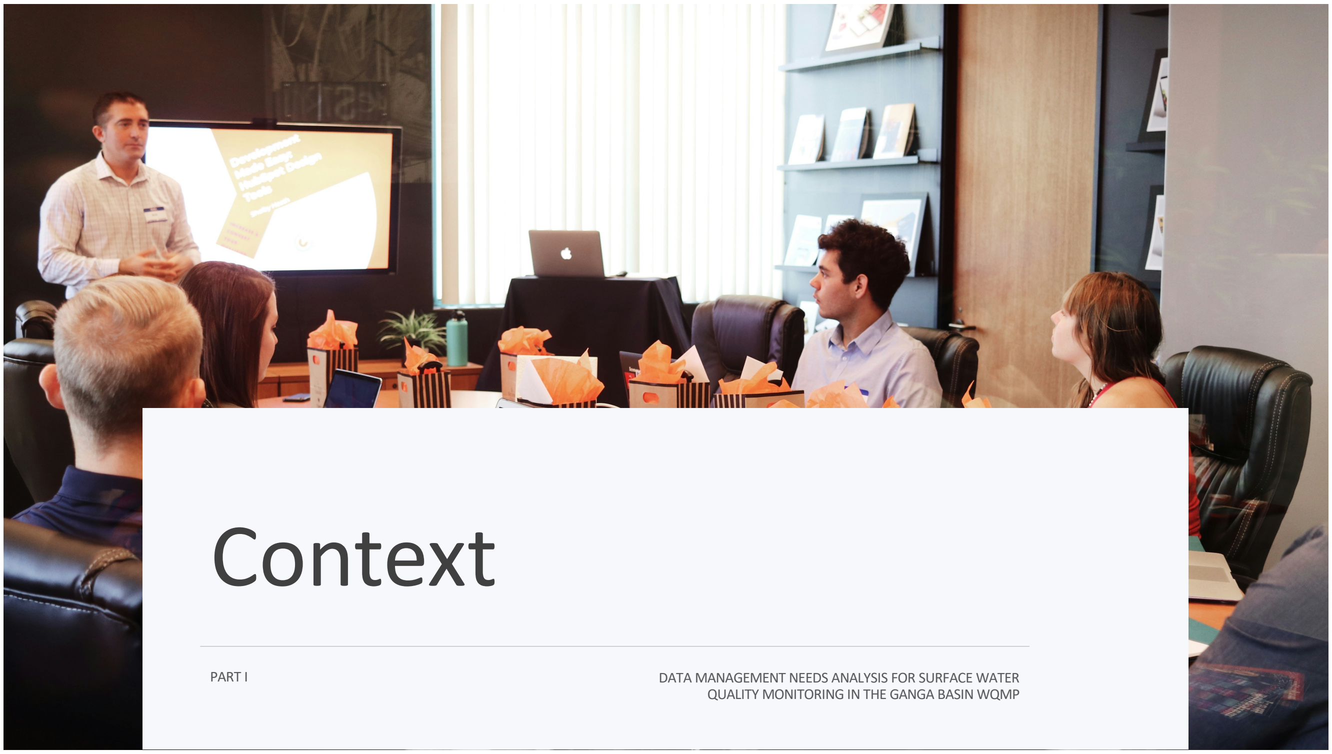
II.1. Information gathering

II.2. Analytical framework

III. Results

III.1. Block 1: Managing WQMP rationale information	III.2. Block 2: Managing field data (sampling context)	III.3. Block 3: Managing materials and processes in regional labs	III.4. Block 4: Data transfer & validation between regional and central labs	III.5. Block 5: Data transfer between CPCB, SPCB, NMCG & SMCG	III.6. Block 6: Data management & communication by UPPCB (A), NMCG & SMCG (B) & CPCB (C)	IV.1. Current situation	V.1. Future situation

OVERVIEW
DATA MANAGEMENT NEEDS ANALYSIS FOR SURFACE WATER QUALITY MONITORING IN THE GANGA BASIN WQMP



Context

PART I

DATA MANAGEMENT NEEDS ANALYSIS FOR SURFACE WATER QUALITY MONITORING IN THE GANGA BASIN WQMP

CPCB	The Central Pollution Control Board (CPCB) coordinates numerous sampling and environmental data collection activities within the Ganga River basin including environmental monitoring of the Ganga, and Ganga tributaries.
Ganga Data	However, the lack of reliable (quality assured) water monitoring data has been recognized as one of the decisive bottlenecks for the development of specific, appropriate, and well targeted pollution-mitigation measures as well as policy instruments for the Ganga river basin.
PTB-NMCG	The project 'Strengthening Quality Infrastructure for Water Monitoring of the Ganges River' was launched in 2018 and renewed in 2023 as a part of the bilateral cooperation between the Government of India and the Federal Republic of Germany. The project is implemented by the National Metrology Institute of Germany (Physikalisch-Technische Bundesanstalt, PTB) in cooperation with the National Mission for Clean Ganga (NMCG)
Standard	To obtain and maintain robust, defendable and comparable data, it is essential to ensure that all the environmental monitoring requirements are met at all stages from planning to the execution, data and metadata management, data sharing, reporting and use of data for implementing environmental regulations and action plans and enforcing existing regulations.
PTB- WSME	PTB has contracted WaterShed Monitoring Europe (WSME), which is specialized in the field of planning, managing and optimizing environmental monitoring, to undertake the analysis of current data and metadata management and data flows for the surface water monitoring data of the Ganga basin.

Methodology

PART II

DATA MANAGEMENT NEEDS ANALYSIS FOR SURFACE WATER QUALITY MONITORING IN THE GANGA BASIN WQMP

What are we aiming to achieve?

Objectives of the mandate:

- Describe the current workflow, analyze needs, and assess communication practices for exchange of river monitoring data.
- Propose targeted workflow improvements and develop recommendations to address identified gaps in data methods .

II.1. Information gathering- Participative and iterative approach January - July 2024

- Four workshops with the key stakeholders of four regional laboratories with a final workshop held in person at NMCG on July 6th.
- Questionnaire survey to understand the current workflows on data management and transfer within the laboratories as well as with CPCB, NMCG and other partners- data management software used by the laboratories and partners.
- Final workshop on data flow between CPCB and NMCG, data management and data use by NMCG.
- Benchmarking workshop with a laboratory from Quebec, Canada, to compare data management workflows and use of Laboratory Information Management Systems (LIMS)/ other databases.
- Meeting with the former Director & Head of the Quality Council of India to validate some elements for data management & documentation required to receive & maintain laboratory accreditation.

This analysis also integrates information obtained during field work observation trips and nine workshops held as part of this project.

II.2. Analytical framework

Current data management practices were analyzed through six blocks:

Block 01

Management of information related to the underlying rationale of the WQMP

Block 02

Managing fieldwork-related data (measurement context).

Block 03

Management of materials and processes in the regional laboratories.

Block 04

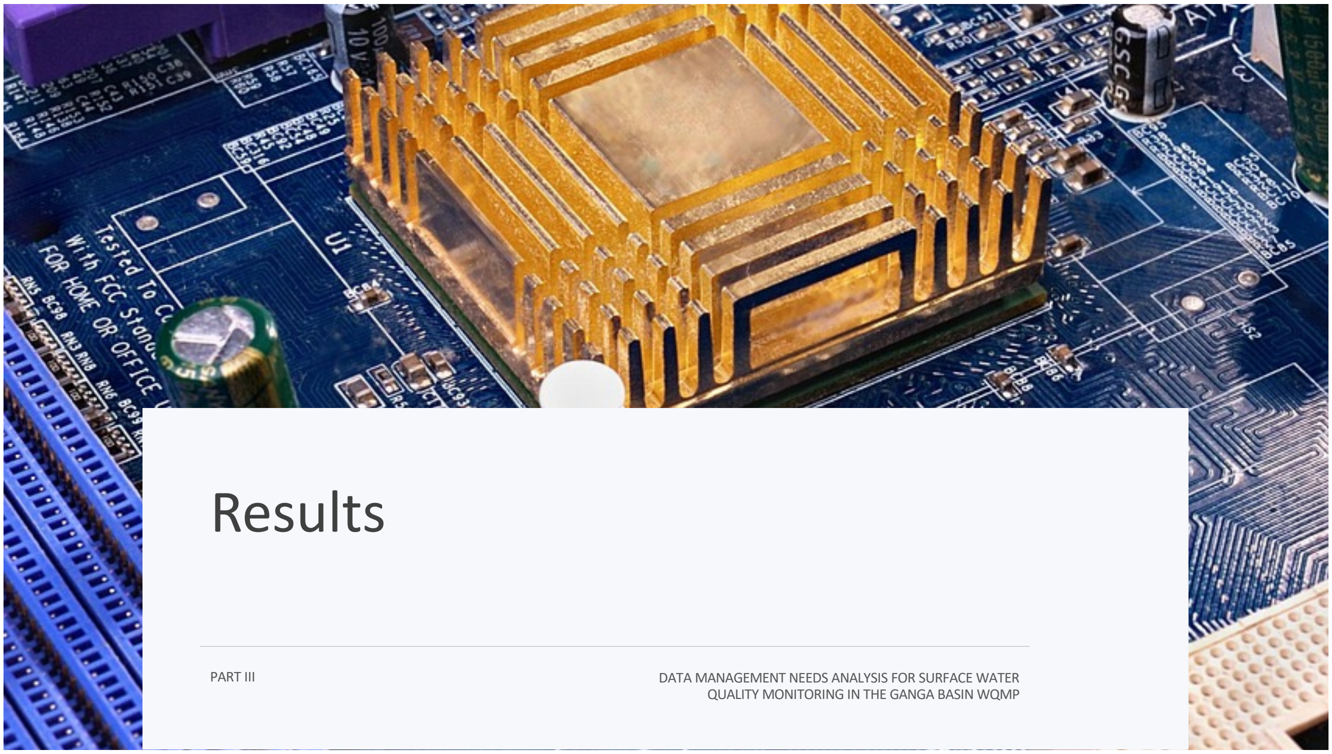
Data transfer processes between the regional laboratories and the central laboratory.

Block 05

Data transfer processes between Central Pollution Control Board (CPCB), State Pollution Control Board (SPCB), & National Mission of Clean Ganga (NMCG)/State Missions of Clean Ganga (SMCGs).

Block 06

Data management & communication by CPCB (Website);
Data management & communication by SPCB (LIMS)
Data management & communication by NMCG / SMCGs (PRAYAG/GIS).



Results

PART III

DATA MANAGEMENT NEEDS ANALYSIS FOR SURFACE WATER QUALITY MONITORING IN THE GANGA BASIN WQMP

III.1. Block 1: Managing WQMP rationale information

Criteria	Status	Notes (based on Fieldwork observations)
WQM objectives	Documented, but not within a data management system	WQM objectives may not be fully aligned with sampling strategies.
Methodology for identifying monitoring objectives and translating them into sampling site selection, water quality parameters, and sampling frequency and recurrence.	Not readily available.	Need for a comprehensive review of the current WQMP, including the sampling site network, site validation, and sampling strategy (frequency, parameters, and sample distribution).
Justification of sampling sites	Information available from field workers and laboratory staff verbally	Need to revise sampling site selection and their justifications, with potential additions or removal of sites, as necessary.
Sampling strategy for each site (how sampling is conducted, where it is conducted and what is sampled)	The information can be made available through documentation transmitted from CPCB to the laboratories	Sampling strategy for each site should be improved and tailored to its specific conditions and objectives.
Sampling methodology (global & site-specific)	The global methodology is available through documentation provided by CPCB to the laboratories. However, the site-specific methodology appears to be unavailable, except for the designated water quality parameters.	While field workers appeared to know where, what, and how to sample at each site, it is unclear how this information is documented. They carried no paperwork to validate their site-specific approach. Fieldwork protocols are too general, lacking site specificity, and are challenging to apply in practice.

III.2. Block 2: Managing field data (sampling context)

Criteria	Status	Notes
Field notes	Yes (paper only & incomplete information)	**Included:** Sampling site name, geographic coordinates, field worker names, date and starting time, water temperature, pH, oxygen titration, general weather conditions, and general notes. **Not included:** Sampling end time, site-specific protocols, sampling objectives, sampling tools and their protocols, standardized weather information (sky type, precipitation, wind, ambient temperature, atmospheric pressure), probes used, probe calibration details, and photos.
Field notes management	Yes, in folders (paper only)	Field notes are not transcribed into any type of data management system, except for pH, water temperature and dissolved oxygen results, which are transcribed into the CPCB LIMS (see block 6)
Field notes use (laboratory)	Yes	Field notes are made available to laboratory personnel
Field notes use (data users)	Yes	Field notes are made available to scientists interpreting the data

III.3. Block 3: Managing materials and processes in regional labs

Criteria / Necessary information to be reported	State	Notes
Maintenance of field materials : Probes/Sensors	Yes	It was not possible to have clear information on field material maintenance procedures or documentation of maintenance results. This information was not clearly managed through a database.
Maintenance of field materials: Sampling tools	Yes	
Maintenance of laboratory instruments	Yes	While procedure documentation is required for laboratory accreditation, this information is only documented on paper, and not through a LIMS, except for laboratories of CPCB (see block 6)
Management of laboratory processes	Yes	
Human resources management	Yes	It is not clear how information on human resources management, especially on (recurrent) training, is documented or obtained.
Quality control & quality assessment, from reception of samples to delivery of results	Yes	Required for laboratory accreditation, but not managed through a LIMS (see block 6)
Quality control & quality assessment of field work, from field work preparation to delivery of sample to the laboratory	Yes	It is not clear how initial sampling protocols are being implemented or how the QC/QA of field work is being enforced and documented. It is clear, however, that none of this information is being documented / tracked using a data management system. NABL accreditation for Central Laboratories but not for Regional Laboratories

III.4. Block 4: Data transfer & validation between regional and central labs

Criteria / Necessary information to be reported	State	Notes
Every laboratory should have access to its own LIMS	No	There is only one central LIMS at the UPPCB level. However, it does not appear to centralize information on the management of each regional laboratory. The extent to which the UPPCB LIMS is used or parameterized remains unclear. Nevertheless, it seems to allow results to be traced back to individual laboratories.
LIMS parameterization should be streamlined, particularly for water quality parameter names, measurement units, methods, maintenance, human resources management, and QC/QA procedures.	Does not apply as not every laboratory has its own LIMS	Each laboratory should digitize the internal processes in their own LIMS in order to be able to transfer final results to a central data base
Results should be transferred through unified electronic transcripts	No	Currently, the analytical results are transcribed manually into the central LIMS at SPCB. However, most field notes, etc. are not transcribed. Manual transcription can result in transcription errors. The information on field conditions is eventually lost (or tracing back to field notes is made complicated).
For a WQMP executed under one central WQMP governance body, a database should be made available that enables management of the underlying rationale of the WQMP, by integrating field data/notes, laboratory results & any other type of environmental data.	No	Such a database does not exist at this point, either at CPCB or at NMCG. Refer to block 6.
Data validation procedures	Yes	Data validation procedures are in place; however, they seem to be fastidious and carried out by email. It is not clear how well this works, or how streamlined the process is.

III.5. Block 5: Data transfer between CPCB, SPCB & NMCG

Criteria / Necessary information to be reported	State	Notes
Data transfer protocol	No	There is no clear data transfer protocol between CPCB and NMCG. Currently, CPCB provides data to NMCG upon request through PDF laboratory reports, which require manual transcription by NMCG staff, increasing the risk of errors. While CPCB can also provide data in Excel format upon request, the specific contents of these files remain unclear. Additionally, CPCB publishes some results on its website, but the data is limited, and no metadata is included. This also applies to SMCG Uttar Pradesh and UPPCB. SPCB transfer data to CPCB through EWQDES portal*.
Unified electronic format including metadata	No	No unified electronic format for data & metadata transfer has been established.
Access to the central database through a specific (limited) user access	No	Since no central database exists (including field data, etc.) access does not exist. In addition, NMCG does not have any type of access to the EWQDES of CPCB.
Data use protocol	No	There appears to be no established data use protocol. Key aspects remain undefined, such as how data should be used, whether results below detection or quantification limits should be reported, which statistical tools should be applied (as different tools yield varying results for the same data), and the limitations of the data in relation to the original objectives.

*CPCB shares water quality data on quarterly basis with National Water Informatics Centre (NWIC) for integration on INDIA-WRIS i.e. Water Resources Information System platform developed by Ministry of Jal Shakti (MoJS) for keeping all water related data on a single platform

III.6. Block 6: Data management & communication by UPPCB (A), NMCG (B) & CPCB (C)

UPPCB (A) Type of system	State	IMPORTANT NOTE: UPPCB is in charge of executing the WQMP (with regional laboratories)
LIMS	Yes	Only for UPPCB—not clear how it is parameterized and how it allows tracking of regional laboratories, apart from opening the portal for the reginal laboratories to enter analytical results manually. Each regional laboratory should have its own specific LIMS, of the same type, and streamlined for parameterization & operation.
GIS	?	Not known to be used for the WQMP data
Project management system	?	Information not available
Website	Yes	Selected data only from the WQMP is shared on the website, without any metadata
Real-time monitoring data cloud	Yes	Real-time data is made available through an online **visualization** platform. However, neither data validation processes nor information on maintenance, calibration, local impacts, etc., are transparent
Integrative field data, metadata, laboratory data and environmental data management system	No	We highly recommend implementation of such a system

III.6. Block 6: Data management & communication by UPPCB (A), NMCG (B) & CPCB (C)

NMCG (B) Type of system	State	Notes
LIMS	No	Not necessary, as NMCG does not operate any laboratories
GIS	Yes	PRAYAG also includes a GIS. However, water quality data in GIS systems are ALWAYS transformed data. Thus, symbols such as < and > and other metadata cannot be managed using a GIS. PRAYAG does not include a data exploitation and visualization platform designed to manage water quality data & metadata. The GIS is not accessible to SMCG's.
Project management system	Yes	PRAYAG includes a project management system. However, it is not able to manage water quality data & metadata.
Website	Yes	Water quality results are made available, but with limited information on metadata and limited possibilities of readily using the data.
Real-time monitoring data cloud	Yes	PRAYAG includes access to the real-time data system from CPCB (see comments above).
Integrative field data, metadata, laboratory data and environmental data management system	No	None of the water quality data from CPCB is being managed in such a system (or any other type of system). Data is used for the GIS system and for reports only. We highly recommend that NMCG interface with (get access to) such a system as implemented by CPCB.

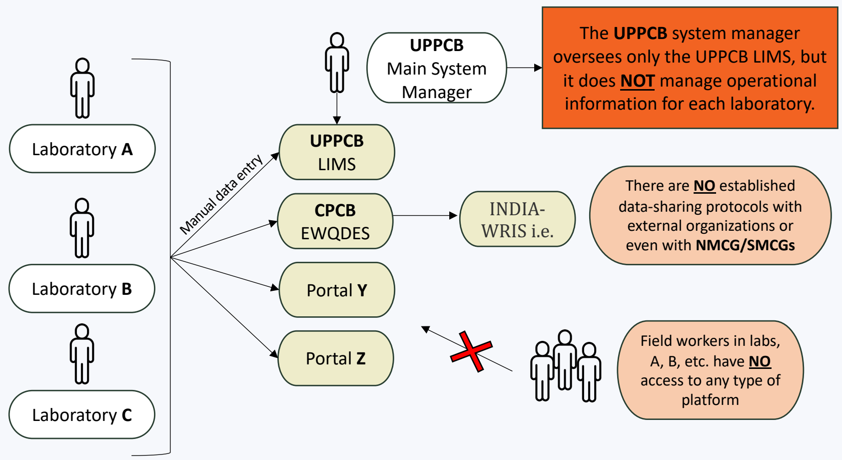
RESULTS DATA MANAGEMENT NEEDS ANALYSIS FOR SURFACE WATER QUALITY MONITORING IN THE GANGA BASIN WQMP

III.6. Block 6: Data management & communication by UPPCB (A), NMCG (B) & CPCB (C)

CPCB (C) Type of system	State	IMPORTANT NOTE: CPCB is in charge of planning the WQMP
LIMS	Yes	LIMS is operational for lab operations, inventory and quality control in CPCB.
GIS	?	Not known to be used for the WQMP data
Project management system	?	Information not available
Website	Yes	Selected data only from the WQMP is shared on the website, without any metadata
Real-time monitoring data cloud	Yes	Real-time data is made available through an online **visualization** platform. However, the data validation processes are not transparent, nor is information on maintenance, calibration, local impacts, etc.
Integrative field data, metadata, laboratory data and environmental data management system	No	We highly recommend implementation of such a system
EWQDES	Yes	Only for CPCB—not clear how it is parameterized and how it allows tracking of regional laboratories, apart from opening the portal for the reginal laboratories to entre analytical results manually. The data should be automatically integrated from the regional laboratories into the EWQDES.

IV.1. Current situation

Visual Representation (current situation)



1. All laboratories receive written work orders from CPCB.
2. No laboratory has its own LIMS.
3. Field workers lack access to the system for work orders, sampling data, probe calibration, or field notes.
4. Regional lab staff manually enter results into UPPCB LIMS and other platforms.
5. Validation/correction requests come via email; only UPPCB LIMS has a clear correction process.

Current situation only includes data-flow to **NMCG** and not to other organizations such as **SMCG** and **SPCB**.
There is also an important time lag on sharing data-based results / directions by CPCB

NMCG has to request data, by email, from **CPCB**. Generally, data is transferred through PDF reports only, unless otherwise requested.

Information from CPCB:

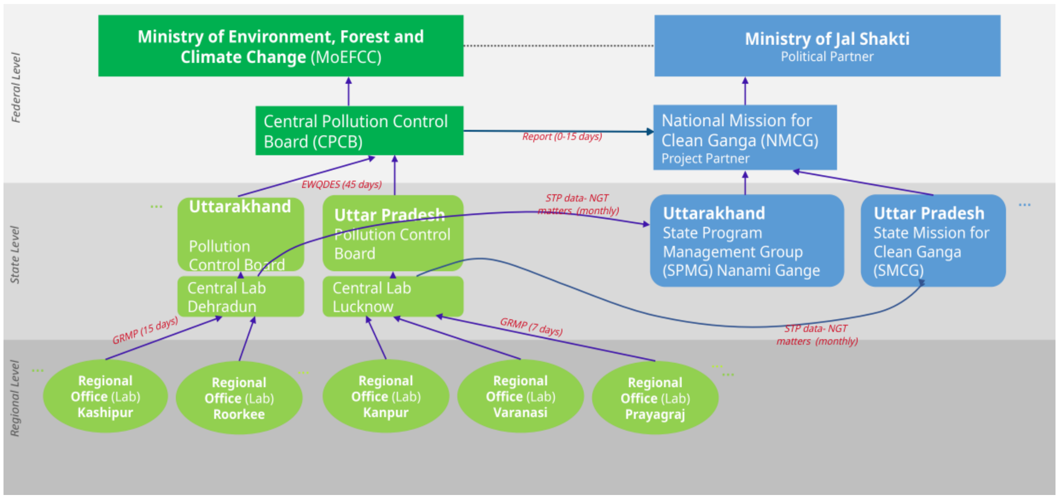
CPCB holds review meetings with all regional offices of SPCBs/PCCs regarding submission of data under NWMP w.r.t data completeness, delay & related quality assurance aspects and provide its minutes of meeting and status report to respective Regional Directorate and CPCB head office.

CPCB regularly reviewed and analysed NWMP data. In case of any outlier or abnormal value, the concerned agencies were informed through various communications (e.g., dated 02.06.2024, 06.04.2021, etc.) to rectify data gaps and ensure the uploading of correct data on the portal.

Workflow description (current situation)

Description
The global WQMP is planned by CPCB. The rationale of the planning decisions is not well documented, except for the objectives. The states have their own State Water Quality Monitoring Programs, with some of the same limitations as the global WQMP.
Regional laboratories receive work orders (sampling sites, water quality parameters, sampling frequency) which they execute according to generic protocols, in addition to other sampling and analytical activities. Sampling routes, and site-specific sampling strategies are not clearly documented.
The field workers' field observations are noted on paper and delivered/submitted to the analytical staff of the laboratories, but much of the information (e.g., meteorological conditions, observations on pollution or changes at the sampling site) is not entered into a database, except for the field results, such as dissolved oxygen.
Analytical results from the laboratories are manually entered into several portals: UPPCB LIMS, CPCB EWQDES, and potentially others.
The individual laboratories do not have a LIMS that would enable them to *manage* their operations.
If there is an error in the results, there is a protocol to correct the data in the UPPCB LIMS. In case of errors in EWQDES, the states raise the request of validation to CPCB by mail. However, it is not clear whether there is a protocol to correct them in all the portals, nor is it clear how the data integrity/coherence in parameterization is overseen for all the portals.
None of the data is automatically transmitted to NMCG.
Sometimes the data is transmitted by PDF, requiring it to be transcribed by hand by NMCG and leading to additional potential transcription errors.
To receive the data by Excel files requires an additional request.
The data is never accompanied by metadata (field work conditions, underlying rationale of the sampling, sampling strategies, or even information on how to use data below the detection limits, etc.). There is a complete absence of data sharing protocols.

IV.1. Current situation – Stakeholder diagram & Data Flow



EWQDES- Environment Water Quality Data Entry System
GRMP- Ganga River Monitoring (Kumbh) Portal
STP- Sewage Treatment Plant
NGT- National Green Tribunal
Note: There are more regional offices – they all follow the same reporting structure

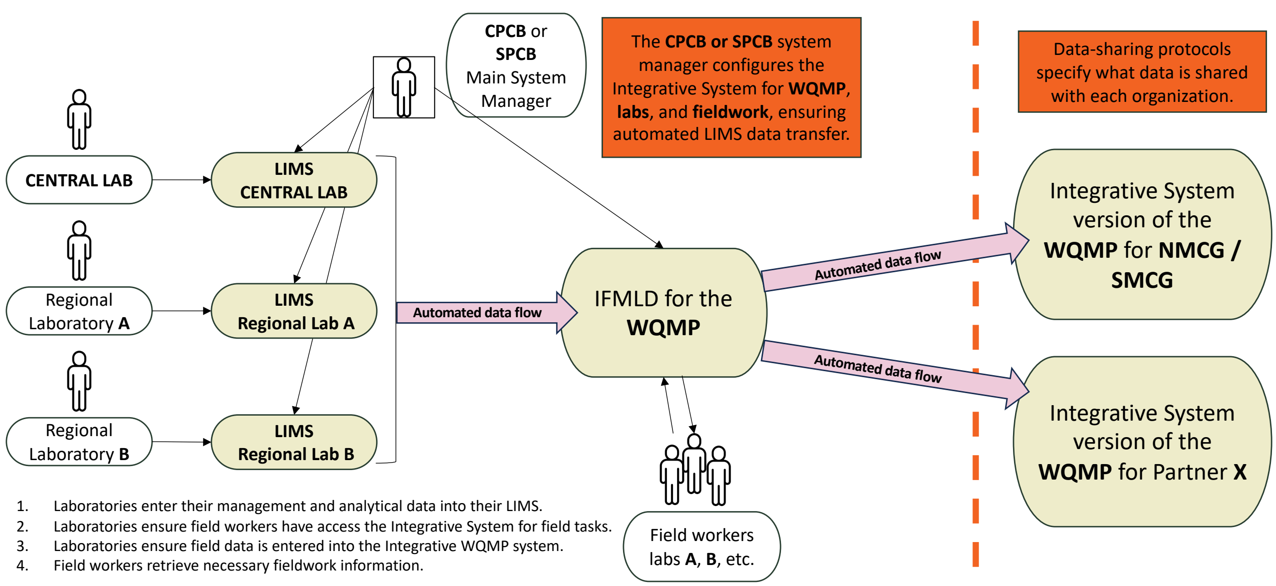
Data Flow
Political Coordination

Noteworthy Observations

Absence of a **Integrative field data, metadata, laboratory data and environmental data management system.**

Absence of a **LIMS for each laboratory to manage and organize data related to laboratory processes.**

Visual Representation (proposed situation)

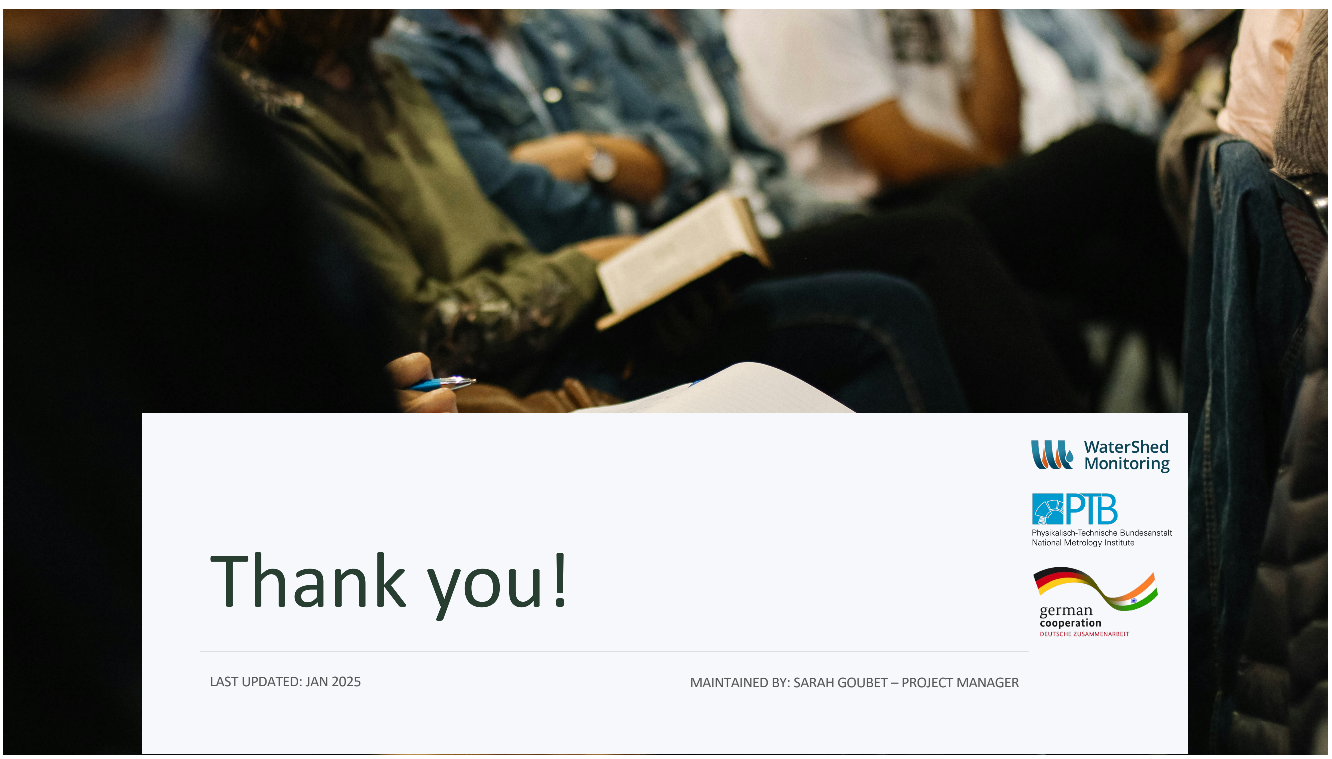


1. Laboratories enter their management and analytical data into their LIMS.
2. Laboratories ensure field workers have access the Integrative System for field tasks.
3. Laboratories ensure field data is entered into the Integrative WQMP system.
4. Field workers retrieve necessary fieldwork information.

Suggested approach – description

Suggested workflow description

- The global WQMP is planned by CPCB and SPCB. The rationale of the planning decisions must be documented and be made transparent, ideally within an **Integrative field data, metadata, laboratory data and environmental data management system.**
- Regional laboratories receive the work orders (sampling sites, water quality parameters, sampling frequency) which they execute according to **site-specific** protocols, in addition to other sampling and analytical activities. Sampling routes, and site-specific sampling strategies **must be** clearly documented.
- The field workers' field observations are noted on paper and delivered to the analytical staff of the laboratories, **and** this information (e.g., meteorological conditions, observations on pollution or changes at the sampling site) has to be entered into the **Integrative field data, metadata, laboratory data and environmental data management system** along with the field results, such as dissolved oxygen, for which the protocols must also be clearly documented.
- The analytical results from the laboratories have to be entered into the individual LIMS of each laboratory enabling them to manage their operations. The results from the LIMS must then be communicated electronically to the **Integrative field data, metadata, laboratory data and environmental data management system.**
- If there is an error in the results, the protocol to correct the results can be found in the **Integrative field data, metadata, laboratory data and environmental data management system.** All the systems must be parameterized on a common basis in order to ensure data integrity & coherence.
- If a correction is needed, the correction must be made in the individual LIMS, and an automated update must be sent to the **Integrative field data, metadata, laboratory data and environmental data management system.**
- **Data sharing protocols need to be established between the stakeholders.**
- The **Integrative field data, metadata, laboratory data and environmental data management system** must provide data according to the rules of data sharing established between the stakeholders.
- The data must be accompanied by metadata (fieldwork conditions, underlying rationale of the sampling, sampling strategies, or even information on how to use the data below the detection limits, etc.) through the **Integrative field data, metadata, laboratory data and environmental data management system.**



Definitions

Database
A database is an organized collection of structured data, typically stored electronically and managed by a database management system (DBMS). This system, along with associated applications, is called a database system or database. Most modern databases model data in tables with rows and columns for efficient processing and querying using SQL.
(Source: What Is a Database | Oracle Canada)

Data
Information, especially facts or numbers, collected to be examined and considered and used to help decision-making, or information in an electronic form that can be stored and used by a computer.
(Source: Cambridge Dictionary online, retrieved 2024-02-19)

Metadata
Information that is given to describe or help you use other information.
(Source: Cambridge Dictionary online, retrieved 2024-02-19)

Real-time monitoring data cloud:
Main purpose of this type of system:
Collects data from on-site hardware, networks, etc. and allows the results to be visualized on user interfaces
(Source: www.techtarget.com)

Definitions

Laboratory Information Management System (LIMS)
Main purpose of this type of system:
Streamlines and optimizes various laboratory processes. Serves as a centralized platform to manage and organize data related to laboratory processes. (Source: Laboratory information management system – Wikipedia)

Geographic Information System:
Main purpose of this type of system:
Stores, analyzes, and visualizes data for geographic positions on Earth's surface. It's a computer-based tool that examines spatial relationships, patterns and trends in geography.

Project management system:
Main purpose of this type of system:
Provides a structured approach to managing operations within a company.

Integrative field data, metadata, laboratory data and environmental data management system
Main purpose of this type of system:
Centralizes all environment-related water monitoring data; including rationale of the WQMP, documentation, field work, sampling context, laboratory results, laboratory tracking (related to information from the LIMS), and metadata. Avoids transcription errors in environmental data, and includes communication features, procedures for streamlined data use, etc. Allows for data publication & sharing under specific conditions & protocols (Source: www.watershedmonitoring.com and Behmel et al., 2019).

Vote of Thanks

We extend our heartfelt gratitude to all the organizations and individuals who have contributed their expertise, support, and dedication to this initiative.

A special thank you to **NMCG-DG, ED-T, Dr. Ishaq, Dr. Shiv Raghuvanshi, and Dr. Hema** for their leadership and guidance.

Our sincere appreciation to the Central and Regional Laboratories of UPPCB and UKPCB for their commitment to water quality management and data integrity.

We also acknowledge the invaluable contributions of SMCG-UP and SPMG-UK, as well as CPCB, for their continuous support in strengthening environmental monitoring systems.

A big thank you to NABL QCI and NPL India for upholding quality standards and ensuring scientific excellence.

Your dedication and collaboration are instrumental in advancing our shared mission. We look forward to continued partnership and progress.

Thank you!

CONCLUSION DATA MANAGEMENT NEEDS ANALYSIS FOR SURFACE WATER QUALITY MONITORING IN THE GANGA BASIN WQMP

Appendix C:

Workflow to establish data sharing protocols with the partners of the WQMP:

1) Identify to whom the information produced by the WQMP is addressed (addressees).
2) Type of information to be transmitted (e.g., Raw data with metadata or pretreated data).
3) Form of information transmission (e.g., providing access to a database, providing a database for sharable data, websites, email) based on existing tools and potential new tools.
4) Identify person(s) in charge of producing information.
5) Identify person(s) in charge of quality control for data or information to be transmitted.
6) Identify transmission timing.
7) Develop a transmission process on an individual case basis.
8) Provide data and metadata, as well as conditions on using each type of data, metadata, and other type of information.

NOTE: Make sure that all QC/QA processes for data acquisition, data management and data validation have been followed correctly before sharing any data, metadata, or information.

The table below is an example of data sharing based on two objectives of the WQMP of River Ganga with different types of data and information sharing:

Table 9: Example of Data Sharing

Monitoring objectives	**Type of information**	**Form of information transmission**	**Person in charge of information production**	**Person(s) in charge of quality control before**	**Transmission timing**	**Examples of Addressees**	**Examples of data sharing conditions**
Identification of state and trends in water quality, both in terms of concentration and effects	Raw data and metadata from the WQMP	Providing access to the IFMLD Excel file with data and metadata (please refer to REPORT EXAMPLE provided with the Laboratory Assessment Grid)	Limnologist of NMCG	Limnologist of NMCG Linguistic revision Members of the scientific reviewer team Director general of the organization	Raw data and metadata: After the data goes through all the QC/QA steps (once a year). Reports: Once a year, time chosen for maximum impact	Raw data: NMCG Limnologist Reports: NMCG staff, CPCB, other government bodies depending on the addresses of the recommendations;	Raw data and metadata: Must not be shared with third parties. Data, metadata, and methodologies must be referenced. Objectives of data acquisition must be reported. Report: Can only be shared after official validation and only to the identified addressees. Credits must be clearly indicated.
Early warning and detection of pollution	Field observations report (note that this process would be different for real time monitoring data)	If nothing special is observed, the field notes are filled out as usual. In the case of observation of a spill, for instance,	Field technician in charge of the field work	Field technician for immediate information (phone call); Laboratory manager for full report	As soon as possible after the observation	Local environnemental agency; police, etc.	A decision support tree with emails and phone-numbers must be established

		pictures must be taken, and a phone call made to a local environmental agency (hierarchy depends on each jurisdiction and stakeholder mapping for these cases)					

Table 10: Most common errors when integrating data from a LIMS into an IFLMD

Error message	Probable causes	Action
Sampling point name not found at line (number of line(s) of the data set to be imported)	Mismatch between LIMS export and IFMLD parametrization (e.g., case sensitive, spelling, empty invisible spaces)	Verify integrity of the data sheet to be imported and correct minor errors if necessary to allow import. If the mismatch is too significant, contact the administrator of the IFMLD who will validate the parametrization of both systems (LIMS and IFMLD)
Sampling context not found	Sampling context not entered. Sampling context not entered for the correct date. Error in the work order of the laboratory (date/time). Sampling context does not contain the right sampling site name. Error on the work order of the laboratory (sampling site name).	Using the field work documentation, validate: Whether the context was entered (if there is a context ID, go to the context ID and correct the entries, if necessary). If there is no context ID, create the context. If contexts are correct, contact laboratory to validate information entered into the LIMS.
Unknown parameter name (followed by the list of parameters)	Parametrization of the LIMS and IFMLD was not standardized. The laboratory has added a new parameter name. The administrator of the IFMLD has deleted or changed the spelling of the parameter name	Validate whether the systems have a mismatch for the names of the same parameter and correct in both systems. Validate whether there is only a slight change in spelling of the same parameter name. In this case, make sure that no new parameter name is added (in case of doubt, refer to SI code).

Unknown units of measurement (followed by the list of units of measurement)	Parametrization of the LIMS and IFMLD was not standardized. The laboratory has added a new unit of measurement. The administrator of the IFMLD has deleted or changed the spelling of the unit of measurement.	Validate whether the systems have a mismatch for names of the same unit of measurement and correct in both systems. Validate whether there is only a slight change in spelling of a same unit of measurement. In this case, make sure that no new unit of measurement is added (in case of doubt, refer to SI code).